



THE FLETCHER SCHOOL

TUFTS UNIVERSITY

Enhancing the Value of Monitoring in International Development: Lessons from Literature and Practice

Jennie Vader
Master of Arts in Law and Diplomacy
2014

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY	3
II. INTRODUCTION.....	5
III. METHODOLOGY	6
IV. THEORETICAL BACKGROUND.....	7
ACCOUNTABILITY AND LEARNING.....	10
THEORETICAL VALUE OF MONITORING	14
V. MONITORING IN PRACTICE.....	22
MONITORING REQUIRES ABUNDANT RESOURCES.....	24
INTRASIGENT TARGETS.....	27
OVERSIMPLIFICATION OF MEASURES.....	30
ANALYSIS OF ACCOUNTABILITY AND LEARNING IN PRACTICE	33
VI. ANALYSIS OF KEY FACTORS INFLUENCING MONITORING.....	38
SUPPORTIVE LITERATURE.....	42
DECENTRALIZED DECISION-MAKING	45
ESTABLISH A LEARNING CULTURE	50
INTEGRATING MONITORING: THROUGH PROCESS AND PEOPLE	54
SUMMARY OF ENHANCING FACTORS	59
VII. CONCLUSION.....	60
ANNEX A: ACKNOWLEDGMENTS	65
BIBLIOGRAPHY.....	66

I. EXECUTIVE SUMMARY

“There is a huge and expensive industry of ‘monitoring and evaluation’, most of the results of which is worth less than a pitcher of spit. We should dismantle it, and use a fraction of the money to fund a smaller, more sharply focused, more rigorous, international, independent collection of real evidence about the cost effectiveness of development interventions.” – Owen Barder¹

This paper seeks to answer the question: Is there empirical evidence that monitoring contributes value to international development programming? In order to find the answer, many more questions, along the lines of the quote above, had to be asked: How much does monitoring *cost*? What *value* does it provide? What is the true *purpose* of monitoring and is it being fulfilled? This paper uses academic theory, practitioner perspectives, and donor policies as evidence to demonstrate that, currently, monitoring is not contributing its full potential to development programming. Some organizations have found significant value in monitoring, but overall, its contribution to the dual purpose of learning and accountability is diminished by a number of systemic factors. The new framework presented in this paper helps analyze the current and potential contribution of monitoring, and provides some solutions which can help enhance its value to development programming. The following are some summary conclusions from this analysis:

- **Little Empirical Evidence:** No large-scale studies were found on the use or contribution of monitoring in international development. Two large-scale, US-based studies were conducted on whether and what type of performance measurement information organizations collected and whether managers used it to inform decisions (performance measurement is used synonymously with monitoring). The studies suggest that US nonprofit organizations may be systematically collecting demographic and short-term data, but the extent to which managers use data is inconclusive. Even though these studies were not in the field of international development, their lessons may be extrapolated to a certain extent as they examined organizations which provided similar services and had similar funding and reporting structures to the programs discussed in this paper.
- **Academic Theory Does Not Actually Discuss the Contribution of Monitoring:** The academic literature presents some benefits of monitoring, including that it can trigger learning cycles, transfer knowledge, and facilitate the formulation of new organizational strategies. However, scholars tend to pay more attention to critiques, such as that monitoring

¹ Barder, Owen. The Lethal Effects of Development Advocacy. Owen Abroad. 2009

uses too many resources, results in oversimplifying the measurement of programs, leads to perverse incentives, and is plagued by logistical challenges such as unclear terminology or the use of irrelevant measures. In particular, most of the challenges cited are not necessarily costs or disadvantages of *monitoring*, but are consequences of the system in which it exists.

- **The Practical Outlook on Monitoring Also Reflects Tension with the System:** Online discussions and conversations with practitioners also generally reflect tensions with what is often called the “results agenda,” and less with monitoring itself. Concerns of many practitioners are that emphasizing the collection of data and use of results to make decisions has led to a narrow focus on outputs and numbers, diminishing the nuance and importance of complex programming and change. Additionally, a funding structure where donors increasingly require data as evidence of programmatic success means that, oftentimes, the type of data collected is externally stipulated and primarily used for upward accountability. This makes many practitioners feel locked in to pre-determined targets, unable to adjust programming, and may make them lose faith in monitoring altogether. Practitioners may then neglect to adequately monitor programs or may subvert the reporting system. Many of these perceptions are contradictory to the messages that donors convey in official M&E policies and documents, revealing a misalignment of expectations and attitudes regarding monitoring between practitioners and donors.
- **The Focus on Upward Accountability Inhibits the Potential of Monitoring:** Without full-scale primary research, there is no conclusive evidence about the extent to which monitoring contributes value to international development programming. However, it is clear that the conflation of monitoring with reporting (upward accountability) diminishes the potential of monitoring to contribute to both accountability and learning. Thus, a new framework places emphasis on the different uses of monitoring for each level of the donor-implementer hierarchy, from participants to field staff to mid-level management and leadership, and finally up to donors. In order for monitoring to contribute to learning and accountability at each level, donors and implementers should decentralize decision-making, establish new incentive systems, foster learning cultures, and integrate monitoring across organizational processes and people.
- **Using the Framework to Enhance the Contribution of Monitoring:** Three case studies demonstrate how decentralized decision-making, incentive systems, learning cultures, and integrated processes can, among other things, increase the use of monitoring towards restructuring programs, better serving target audiences, and utilizing resources more efficiently. Recognizing that the entire structure of international aid will not change drastically in the near future, these case studies highlight how the necessary conditions can be created to help enhance the value of monitoring.

II. INTRODUCTION

Conventional wisdom indicates that significant time and attention is being paid to Monitoring and Evaluation (M&E) in international development, but is there any empirical evidence that M&E is contributing value to development programming? A cursory search for the value or use of M&E reveals substantial amounts of theory and empirical studies on evaluation use and utilization. However, there is a significant gap in the literature concerning monitoring. Most of the publicly accessible information consists of how-to guides and organizations' published quarterly or annual reports, but there are few empirical studies demonstrating whether and how monitoring is actually used or valued. Therefore, this paper applies some lessons from related bodies of literature such as Evaluation Use, Performance Management and Nonprofit Effectiveness, in addition to current perspectives from the field, to examine whether, and if so how, monitoring is contributing value to international development programs. The relevant literature and methodology used will be featured in the analysis of each section.

The criteria used to determine whether monitoring contributes value to international development programming is whether monitoring information is used in both accountability and learning. Each element is necessary, but not sufficient in its own right, to fulfill the purpose of monitoring. However, different organizations may vary in the amount that accountability and learning are fulfilled; in a highly effective organization, monitoring would contribute to upward, downward, and internal accountability while also informing decisions and positively contributing to learning in the organization. In order to analyze the potential value of monitoring, this paper first explores theoretical contributions and challenges of monitoring, and then applies those theories to experiences of practitioners. These perspectives culminate in a framework which helps analyze the value of monitoring at each step of the donor-implementer hierarchy. This framework, presented on page 39, demonstrates the discrepancy between monitoring's current and potential contributions as well as possible ways to rectify those issues which currently diminish the value of monitoring. Finally, case studies help elucidate the factors needed to enhance the contribution of monitoring.

Currently, there is evidence that the attitudes and expectations of monitoring vary between donor agencies and practitioners. The power dynamic created by top-down funding from donors to implementing organizations creates an environment where misaligned

expectations result in an unbalanced focus on upward accountability. Thus, the potential value of monitoring is diminished by undermining other key purposes such as downward accountability and learning. By recognizing that many different levels of practitioners may come into contact with and utilize monitoring differently, organizations can alter the type of information gathered and the systems in place to analyze and use that information. Factors that can help enhance the value of monitoring to various stakeholders include decentralized decision-making power, new incentive systems, an organizational culture centered around learning, and integration of monitoring across organizational processes and people.

III. METHODOLOGY

This paper primarily uses information from scholarly work in the fields of Monitoring and Evaluation, Evaluation Use, Performance Management, Public Administration, and Nonprofit Effectiveness. This literature was accessed through a direct search and by mining relevant sources from prominent pieces of literature. A significant amount of information was also taken from Monitoring and Evaluation email listservs, discussion communities, and blogs, to get the most current perceptions of practitioners in the field. The information from these online resources was supplemented by informal conversations with seven M&E practitioners who willingly responded to a request on the Pelican email listserv; a list of those practitioners can be found in Annex A. Finally, publicly available documents from the top ten aid-contributing countries of the Organization for Economic Cooperation and Development (OECD) were analyzed based on relevant M&E issues and the specific methodology is provided in Section V.²

For the purposes of this paper, information was considered evidence if it showed a broad consensus along a variety of sources. Although most of the information for this paper is a collection of practitioner perspectives, it represents a wide range of viewpoints, many of which are substantiated by academic research from several different bodies of literature. No one ‘expert

² Statistics found at <http://www.oecd.org/statistics/>. Agencies that donated the most bilateral official development assistance in 2011 according to the DAC: US: USAID usaid.gov; Germany: German Organization for International Cooperation (GIZ) giz.de; France: French Agency for Development (AFD) afd.fr; UK: Department for International Development (DFID) www.gov.uk/government/organisations/department-for-international-development; Japan: Japan International Cooperation Agency (JICA) jica.go.jp; Netherlands: Government of the Netherlands government.nl/issues/development-cooperation; Australia: AusAid ausaid.gov.au; Canada: Foreign Affairs, Trade, and Development (CIDA) acid-cida.gc.ca; Norway: Norwegian Agency for Development Cooperation (NORAD) norad.no; Sweden: Swedish International Development Cooperation Agency (SIDA) sida.se.

opinion' was given priority over other data, but the analysis could certainly be strengthened by more opinions of field staff and local participants. A more detailed definition of evidence and how it relates to monitoring is provided below.

The scope of this study is significantly limited by time and resources which precluded systematic collection of primary data. A large-scale study could help better elucidate the multiple realities and perceptions of whether monitoring has contributed to the field of international development. While much work has gone into addressing the multi-faceted nature of the topic, the scope of this study is certainly a limitation to a comprehensive analysis.

IV. THEORETICAL BACKGROUND

Monitoring in international development is defined here as the ongoing collection and analysis of data on the progress towards results and changes in the context, design, and implementation of programs. There are many definitions and different perceptions of monitoring which complicate an analysis of its true value. Therefore, this section will first describe some good practices taken from three commonly cited manuals.³ The list of good practices below is by no means a comprehensive list, but instead constitutes the minimum and most widely agreed-upon characteristics of an efficient and effective monitoring system as it should ideally operate. The theoretical descriptions and current monitoring practices discussed in the rest of this analysis will be examined against these principles of "good monitoring".

- **Planning:** The intended use of monitoring and the monitoring plan should be created during the program design phase. Indicators and targets should be designed after establishing clear goals, objectives, activities, and outputs.⁴ Indicators are quantitative or qualitative variables that signal the changes sought by a program; they should be formulated from the goal and objectives and are the basis of a monitoring system. However, it is important that the system

³ A. Causton et al., "Guidance on Monitoring and Evaluation," *Catholic Relief Services* (October 2012).; Sonia Herrero, *Integrated Monitoring: A Practical Manual for Organisations that Want to Achieve Results in Progress* UG,[April 2012].; C. Church and M. Rogers, *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*. Search for Common Ground, March 2006.

⁴ Definitions of terms used from DfR: **goal**- the broadest change that the program seeks to achieve, the highest order change; **objective**- types of change that are needed in order to achieve the goal, sub-order or intermediate changes; **output**- deliverables or products resulting from a program's activities; **target**- the size or magnitude of the intended change

is light-touch, meaning only information that will inform a specific decision should be collected.⁵ This reduces the amount of unused data, or “noise,” and allows for a responsive system that produces timely data.

- **Type of Information:** Monitoring information should be collected on project implementation, answering the questions: are activities being conducted as planned, are they meeting quality standards, are the number of outputs reaching set targets? Monitoring information should also be collected on progress toward results – whether objectives are being met. Both monitoring on project implementation and progress towards results are important to fully understand whether the program is being conducted as planned and catalyzing the expected changes. A good monitoring system also allows organizations to test the assumptions that are built into the program and important aspects of the surrounding context.
- **Quality:** Data must be valid and reliable, which involves using appropriate tools and methodologies for data collection, well-trained data collectors and inputters, and requires quality checks throughout the process.⁶ Moreover, data should be disaggregated by sex and/or other dimensions that may influence individuals’ situations or perspectives differently.⁷
- **Use:** Raw data must be analyzed and transformed into information which meets the needs of and informs decision-makers. Challenges, weaknesses, and lessons learned should all be identified along with successes and accomplishments.⁸ Finally, project staff and managers should reflect on monitoring information, and the decisions it informs, regularly and in a timely manner. Actions should be adjusted as appropriate, based upon new knowledge and learning.⁹ Also, findings should be disseminated, as appropriate, to multiple stakeholders for their learning and for transparency.¹⁰

⁵ Causton et al., *Guidance on Monitoring and Evaluation*, 4; Church and Rogers, *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs.*, 83 Ibid., 83

⁶ Causton et al., *Guidance on Monitoring and Evaluation*, 4.

⁷ Ibid., 4 and Church and Rogers, *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs.*, 89.

⁸ Causton et al., *Guidance on Monitoring and Evaluation*, 4.

⁹ Herrero, *Integrated Monitoring: A Practical Manual for Organisations that Want to Achieve Results*

¹⁰ Ibid., 4

Implicit in this operational definition of monitoring is that the data collected will be processed, some learning will occur, and then some action will be taken or some decision will be made. Thus, monitoring may be one part of evidence-based decision-making, provided the monitoring information used is considered evidence. The Oxford dictionary defines evidence as “the available body of facts or information indicating whether a belief or proposition is true or valid.”¹¹ Yet, there are many important characteristics which determine whether said belief or proposition is true or valid: the quality of facts or information, the credibility of the information source, the depth and breadth of information available, and the standards to which that information is held. Currently in international development, many practitioners think that the determinant of evidence is methodological appropriateness. This point of view often holds randomized control trials as a gold standard when proving validity.¹² Alternatively, Donaldson presents a framework where evidence is credible when stakeholders perceive it to be trustworthy and relevant for answering their questions. He argues that credible evidence is contingent on the questions at hand and the choices made by stakeholders in light of practical, time, and resource constraints.¹³

Donaldson’s definition of evidence is very useful in terms of monitoring activities and the use of evidence to make decisions. Individual organizations should establish the processes and standards by which information is considered evidence and at what point that evidence may be used to make decisions or take action. For most organizations, it seems unlikely that the standard for good evidence is that which comes from a randomized control trial or other experimental design, but information should also probably not be considered evidence if presented by one or two participants out of thousands or by one expert opinion with no grounding in the specific programming at hand. Ideally, proper data collection and storage processes would be in place so that a decision-maker can assess the ubiquity of a certain piece of information, its data source(s), and relevance of programming in order to determine if that information is reliable and credible enough to warrant action.

¹¹ <http://www.oxforddictionaries.com/definition/english/evidence>

¹² Donaldson, Stewart. "A Practitioner's Guide for Gathering Credible Evidence in the Evidence-Based Global Society." Chap. Epilogue, In What Counts as Credible Evidence in Applied Research and Evaluation Practice?, edited by Donaldson, S., C. Christie and M. Mark, 239. California: Sage Publications, Inc., 2009, 240.

¹³ Ibid., 244-245.

Accountability and Learning

The two main purposes of monitoring are learning and accountability and these are the criteria upon which this analysis assesses the value of monitoring. This section details the theoretical underpinnings of accountability and learning.

Accountability, as defined by Ebrahim, entails four components: transparency, answerability, compliance, and enforcement. Transparency involves collecting information and making it available and accessible to public scrutiny; answerability requires providing clear reasoning for actions and decisions; and compliance means that findings from monitoring and evaluation are transparent. Finally, enforcement or sanctions are put in place for shortfalls in compliance, justification or transparency.¹⁴ The extent to which an organization is accountable and its accountability mechanisms are influenced by many factors: to whom an organization is accountable, for what information, and what repercussions ensue for failure to comply or deliver.¹⁵

Ebrahim defines three types of accountability, characterized by different stakeholders which can hold organizations to account. First, organizations may be accountable to donors, foundations, and governments through what is often called upward accountability. Organizations also often have an internal reporting hierarchy, or internal accountability. Finally, and less well-recognized, is downward accountability, understood to be the relationship between nonprofit organizations and the people to whom they provide goods or services.¹⁶ Murtaza calls this 360-degree accountability and adds sideward accountability to the model, which means organizations are accountable to peer relationships on a professional level.¹⁷ Each of these types of accountability and their audiences may require different types of information about finances, governance, and/or performance, and whether or not the organization is achieving its mission.

Accountability, however, may also be a “means by which organizations and individuals take internal responsibility for shaping their organizational mission and values, for opening

¹⁴ Alnoor Ebrahim, "The Many Faces of Nonprofit Accountability," *Working Paper, Harvard Business School* 10-069 (2010), 3.

¹⁵ Alnoor Ebrahim, "Accountability in Practice: Mechanisms for NGOs," *World Development* 31, no. 5 (2003), 814-15; J. Lee, "NGO Accountability: Rights and Responsibilities," *CASIN: Programme on NGOs and Civil Society* (2004).

¹⁶ Ibid. 814-815. Alnoor Ebrahim, "Accountability in Practice: Mechanisms for NGOs," *World Development* 31, no. 5 (2003), 814-815.

¹⁷ Niaz Murtaza, "Putting Lasts First: The Case for Community-Focused and Peer-Managed NGO Accountability Mechanisms," *International Society for Third-Sector Research* 23 (2011), 113.

themselves up to public or external scrutiny and for assessing performance in relation to goals.”¹⁸ Ebrahim argues that accountability to an organization’s mission specifically embraces a long-term view of performance measurement by emphasizing iteration and learning. This means that an organization should take care to internalize its mission, monitor performance against it, and periodically review the mission in light of changing context. Moreover, goals and strategies should be subject to adaptation.¹⁹

Here, the nexus between accountability and learning becomes more apparent. Accountability is more than just providing information, it is a full cycle that entails taking action. That action may be a result of learning, monitoring’s other purpose. Learning implies the creation of deeper understanding and insight, which goes beyond intake and assimilation of information and knowledge. A common model, introduced by Argyris and Schon is that of single and double-loop learning. Single-loop learning applies to immediately observable processes and structures which may lead to adaptation, but only within existing organizational frameworks. The common metaphor for single-loop learning is one of a thermostat that learns when it is too hot or too cold and adjusts the heat accordingly. Double-loop learning, on the other hand, involves questioning an organization’s underlying norms, policies, and objectives.²⁰ It may at first seem that monitoring is only a single-loop learning exercise where a programmatic adjustment is made to improve effectiveness or efficiency. For example, Britton states,

It is important for NGO staff to recognise that learning is not just about developing innovative new programme designs or policies, important though these are. Learning is equally importantly about what may seem like more mundane matters, for example identifying the small but important improvements in project effectiveness that come about by paying careful attention to the data from project monitoring. In this sense, learning provides a real purpose for gathering monitoring data – indeed many would argue that learning directed at creating immediate improvements to project implementation constitutes the most important purpose for monitoring.²¹

¹⁸ Ebrahim, *Accountability in Practice: Mechanisms for NGOs*, 815Ebrahim, *Accountability in Practice: Mechanisms for NGOs*, 815

¹⁹ Ebrahim, *The Many Faces of Nonprofit Accountability*, 1Ebrahim, *The Many Faces of Nonprofit Accountability*, 1

²⁰ Katherine Pasteur, *Learning for Development: A Literature Review* (No. 6: Brighton: Institute of Development Studies,[2004]). and Britton, *Organisational Learning in NGOs: Creating the Motive, Means and Opportunity*, 41

²¹ Britton, *Organisational Learning in NGOs: Creating the Motive, Means and Opportunity*, 10Ibid., 10

However, by definition, monitoring could also inform foundational assumptions, which constitutes double-loop learning.

The Experiential Learning Cycle, established by David Kolb, is another prominent theory which indicates that individuals learn in a continuous cycle of four stages: doing, reflecting, connecting and testing.²² Thus, learning involves taking an action, reflecting on the outcomes of the action, making connections with what was previously known and understood, and then testing those connections and new ideas into action.²³ A related model is the Knowledge Hierarchy in which unfiltered data are turned into knowledge by placing the information in a framework and giving the data context. People interpret and internalize the information and then make decisions about their behavior based on that knowledge. Wisdom, the final step, is achieved when values and commitment guide behavior.²⁴ Both the Experiential Learning Cycle and Knowledge Hierarchy are similar in that they are processes by which individuals take data and process it into something more – knowledge, behavior change, wisdom.

The way people gain this information may be tacit or explicit. Explicit knowledge can be expressed in words or numbers and can be shared between people easily either in written or verbal form. Tacit knowledge, however, is harder to formalize, communicate, and share with others. It may be understood as ‘know-how’ and often consists of beliefs, ideals, values, and mental models deeply ingrained in individuals.²⁵ These types of knowledge are important because in order for an organization as a whole to be a learning organization, individuals will need to share what they have learned.

Learning as an organization requires collective learning, which means information and wisdom must be articulated and made available to others.²⁶ A useful model for considering how individual knowledge can become collective learning is the Knowledge Creation Spiral, introduced by Nonaka and Takeuchi. The model involves four stages: socialization, externalization, combination, and internalization. First, individuals share tacit knowledge by being involved in joint activities, then externalization transforms individual knowledge to

²² Ibid., 40

²³ Bruce Britton, "Learning for Change: Principles and Practices of Learning Organizations," *Swedish Mission Council*, no. 5 (2002), 7.Bruce Britton, "Learning for Change: Principles and Practices of Learning Organizations," *Swedish Mission Council*, no. 5 (2002), 7.

²⁴ Britton, *Organisational Learning in NGOs: Creating the Motive, Means and Opportunity*, 42

²⁵ Ibid., 43

²⁶ Britton, *Learning for Change: Principles and Practices of Learning Organizations*, 9

explicit collective knowledge by interacting with others. Communities of practice and storytelling, for instance, are said to be useful activities for externalization. Combination utilizes that knowledge through an organization's systems; for example, research or evaluation findings could be used to develop internal guidelines or procedures. Finally, internalization makes explicit knowledge tacit so that it becomes part of an organization's shared mental models and culture.²⁷ This leads to the idea of a learning organization. A related model, the Eight Function Model, suggests that in order to learn effectively, a nonprofit organization must attend to eight functions, including gathering experience, accessing external learning, establishing communication systems, drawing conclusions, developing organizational memory, integrating learning, applying the learning, and creating a supportive culture. This model assumes that learning can only be said to occur when it results in action; central to organizational learning is thus the application of learning to action.²⁸

One can see how monitoring information fits into these models. First, data may be collected formally or informally, leading to either explicit knowledge that is easily shared, or tacit knowledge which, while it may be crucial, is more difficult to communicate. Monitoring could be involved in an individual's learning process if they are in the position to take action, reflect on the outcomes and connections and then test whether an adjustment to programming has an effect (Experiential Learning Cycle). Moreover, one can imagine that monitoring information would be most useful to an organization if the individual has the support and encouragement to enable him/her to articulate the wisdom in a way that is accessible to others (Knowledge Hierarchy). Finally, an organization must have set processes and activities in order to transform the data gathered through monitoring from tacit knowledge into explicit knowledge and individual learning into collective learning (Knowledge Creation Spiral and Eight Function Model).

These accountability and learning frameworks represent the ideal ways monitoring could contribute value to international development programming. The following two sections will explore whether and how they are actually used.

²⁷ Britton, *Organisational Learning in NGOs: Creating the Motive, Means and Opportunity*, 45-46

²⁸ Ibid., 42

Theoretical Value of Monitoring

Due to the dearth of theoretical work on monitoring as an academic construct, this paper will draw heavily on the Performance Management literature. Performance management, also referred to as results-based management, is a broad strategy generally found in public administration and some for-profit companies. It is concerned with the decision-making processes, procedures, and incentives in organizations and aims at achieving results and continuous improvements.²⁹ Thus, in this paper, performance management may be equated with a holistic monitoring system. Consequently, performance *measurement* will be used synonymously with monitoring.

Also, it is important to be clear on other terms which will be frequently used; “donor” and “funder” will be used synonymously for those organizations which donate money to implementing organizations running programs. Implementing organizations are viewed as separate from partner or local organizations, which are locally-based in the host country. Thus, implementing organizations are generally those international organizations in charge of applying for funding, coordinating programs, and liaising with donors. Finally, “participants” will be the term used for the target population of a given program; however, some quotes and other sources also use the term beneficiaries, so that term may be used accordingly.

The concept of collecting information on the progress of social betterment programs began primarily in North America in the last half of the 20th century as the nonprofit sector grew larger and expanded its range of stakeholders. According to LeRoux and Wright, little thought was given to measuring performance in the early years of American charity and philanthropy because oftentimes nonprofit organizations were trusted simply due to homogenous religious and social values.³⁰ Then, issues of accountability and performance measurement began to grow, first in Canada in the mid-1970s and later in the US, during the 1990s.³¹ Some scholars attribute this trend to the fact that nonprofits began playing a larger role in providing public services, which forced governments to account for the spending of those public funds.³² Also, there have been a

²⁹ Annette Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience* (France: DAC Working Party on Aid Evaluation,[2000]).

³⁰ K. LeRoux and N. Wright, "Does Performance Measurement Improve Strategic Decision Making? Findings from a National Survey of Nonprofit Social Service Agencies," *Nonprofit and Voluntary Sector Quarterly* 39, no. 4 (2010), 572.

³¹ Jerome Winston, "Performance Indicators - Promises Unmet: A Response to Perrin," *American Journal of Evaluation* 20, no. 1 (1999), 95..

³² Carman, *Nonprofits, Funders, and Evaluation: Accountability in Action*, 375

number of financial scandals, such as those with the United Way of America and the American Red Cross, which caused a crisis of accountability, eroding public confidence and trust in nonprofits.³³ Another possible reason is that the increasing commercialization within the nonprofit sector and increasing competition between for-profit and nonprofit organizations have caused people to question the legitimacy of nonprofits.³⁴ Specifically in the field of international development, some argue that the increased pressure to measure performance is fueled by 'aid fatigue' – the public's perception that aid programs are failing to produce significant development results, coupled with declining aid budgets. Also, since many governments, particularly in OECD countries, are under pressure from their constituents to account for money spent on development, the agencies that those governments fund are also under pressure to show their results and use of resources.³⁵

Murtaza points out several similar trends in terms of increased calls for NGO accountability. She argues that accountability has become increasingly important to demonstrate that NGOs actually represent the people's interests, to allow greater scrutiny in an effort to mitigate opportunities for NGOs to become 'fundraising covers', to compensate for several high profile cases of abuse of power by well-established NGOs, and because of the increased number of evaluations that questioned the quality of NGO projects.³⁶ Now, the trend towards performance measurement can be seen through many avenues: in the 1993 US Government Performance and Results Act, in results-based management systems established in Canada and many other Development Cooperation Agency (DAC) countries, and in the Managing for Results initiative related to the Millennium Development Goals, as well as in similar performance measurement tools and systems.³⁷

Far more studies have been completed on measuring the success and usefulness of performance management in western countries than have been conducted on monitoring in

³³ K. LeRoux and N. Wright, "Does Performance Measurement Improve Strategic Decision Making? Findings from a National Survey of Nonprofit Social Service Agencies," *Nonprofit and Voluntary Sector Quarterly* 39, no. 4 (2010), 572.

³⁴ Carman, *Nonprofits, Funders, and Evaluation: Accountability in Action*, 375

³⁵ Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*, 19.

³⁶ Murtaza, *Putting Lasts First: The Case for Community-Focused and Peer-Managed NGO Accountability Mechanisms*, 111-112

³⁷ Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*; R. Lahey and S. Nielsen, "Rethinking the Relationship among Monitoring, Evaluation, and Results-Based Management: Observations from Canada." *New Directions for Evaluation* 137 (2013), 45. and Joanne Carman, "Nonprofits, Funders, and Evaluation: Accountability in Action," *The American Review of Public Administration* 39 (2009), 374.

international development. The studies detailed here are those which examine performance measurement, as defined above, and ones that were conducted in nonprofit organizations or donor agencies, but not those conducted in national government structures. A number of studies, chiefly on government sector performance management, were published in the early 2000s, likely after the swell of attention given to the topic in the 1990s; however, the definition and implementation of performance management in these studies are too disparate for use in this paper. Thus, there are two relevant large-scale studies of performance management in nonprofit organizations.

First, Morley, Vincent, and Hatry (2001) provide a broad view of the state of monitoring in US human services organizations.³⁸ The study examined 36 organizations which responded to Independent Sector's 1998 Measures Survey, particularly those that reported collecting outcome information regularly and worked in human services, health and mental health services, as well as those that worked in environmental and animal protection. The researchers found that 83% of the nonprofit organizations regularly collected and tabulated data on at least some outcomes related to results, primarily shortly after completion of services, but some in regular intervals even after services were completed. Most organizations did not use sophisticated data collection techniques: only two or three organizations used sampling procedures and comparison groups while most simply compared outcomes over time. None of the organizations disaggregated the data by demographic characteristics. Approximately half of the organizations reported analyzing their data in order to aid in making program improvements and reporting results, although the specific improvements and results were not given. The audiences for outcome reports were usually organizational boards and funders, including government agencies; additionally, a common use of outcome information was fundraising.

LeRoux and Wright (2010) also studied whether and to what extent performance measurement is used by nonprofit managers to make strategic decisions. These researchers questioned the foundational assumption that managers will use performance information to make purposeful decisions about how to manage programs or allocate resources.³⁹ The large-scale study received responses from 314 executive directors of US nonprofit social service and

³⁸ H. Hatry, E. Morley and E. Vinson, *Outcome Measurement in Nonprofit Organizations: Current Practices and Recommendations*. Independent Sector,[2001].

³⁹ LeRoux and Wright, *Does Performance Measurement Improve Strategic Decision Making? Findings from a National Survey of Nonprofit Social Service Agencies*, 575LeRoux and Wright, *Does Performance Measurement Improve Strategic Decision Making? Findings from a National Survey of Nonprofit Social Service Agencies*, 575

community development organizations. The primary independent variable of interest was the organization's reliance on a range of performance measures in the form of an index consisting of workload and output, unit cost and efficiency measures, outcome and effectiveness measures, client or customer satisfaction, external audits, and industry standards and benchmarks.⁴⁰ Other independent variables were also tested, including professionalism, functional training of the executive director, governance, competition, revenue portfolio, and organization age.⁴¹

The findings indicate that nonprofit managers do appear to use performance information in making strategic decisions, although a strict definition or examples of "strategic decisions" were never provided. LeRoux and Wright report that for each unit increase in the performance measures index, effectiveness of strategic decision-making is expected to increase by .04 standard deviations with high statistical significance.⁴² Other factors that were shown to promote strategic decision-making were effective governance, the executive director's education level, and funding diversity. Their good governance index suggests that organizations whose board provides sufficient direction, makes clear decisions, and has a good working relationship with the executive director will be more effective at making strategic decisions.⁴³ Also, LeRoux and Wright argue that organizations with diverse funding may make more effective decisions because "they have a higher degree of autonomy, the freedom to take risks, and the ability to make decisions based on the context of the 'big picture'."⁴⁴ A competitive environment was shown to have a harmful effect on decision-making likely due to the fact that high levels of competition encourage managers to focus more on the present, resulting in resources being devoted to daily activities such as providing quality services, rather than long-term strategy.⁴⁵

Both of these empirical studies, limited to their respective samples, suggest that nonprofit organizations may be systematically collecting data, but the extent to which managers use data is inconclusive. At best, these studies suggest that demographic and short-term data are provided to organizations' boards, which may or may not use the information depending on a number of other factors. Even though these studies were not in the field of international development, their lessons may be extrapolated, to a certain extent, because they examined organizations that

⁴⁰ Ibid., 578

⁴¹ Ibid., 578-579

⁴² Ibid., 580

⁴³ Ibid., 579

⁴⁴ Ibid., 579

⁴⁵ Ibid., 579

provided similar services and rely on similar funding and reporting structures. However, both studies are limited due to their reliance on self-reported data, which could have resulted in an overestimation of how much organizations actually collect, analyze, and use performance data. Moreover, a sample size of two studies is clearly too small to generalize conclusions about the utilization of performance information and certainly presents issues in translating conclusions to organizations operating in other cultures with different opportunities and constraints. Still, the overall lack of other conclusive evidence on how and whether monitoring is used domestically or internationally, makes these studies at least a starting point for analyzing the contribution of monitoring to international development.

There is, however, other academic work discussing the theoretical value of monitoring. This body of literature is not necessarily based on large-scale empirical studies, but on reviews of relevant literature and the professional experience of the authors. Whether or not the challenges and opportunities materialize will be discussed in the next section; this section simply presents an overview of current theoretical understandings of monitoring and performance measurement. Most of the literature focuses on challenges or costs of monitoring which can be placed into roughly four categories: resources, perverse incentives, oversimplified measures, and logistical challenges.

The resources required for monitoring is a commonly cited challenge. Some costs incurred include the time, expertise, and other inputs needed to collect, analyze, use, and disseminate data.⁴⁶ The costs of monitoring varies by size of the program, amount of data collected, rigor required, etc.; however, they may not necessarily cause undue financial burden depending on the system design. Bamberger's observations of M&E practitioners in Asia reflect common complaints about the time spent on monitoring:

[Conference] practitioners from various countries reported that most of their staff are located in the central office and their time is spent in requesting monitoring data,

⁴⁶Michael Bamberger, "The Monitoring and Evaluation of Public Sector Programs in Asia: Why are Development Programs Monitored but Not Evaluated?" *Evaluation Review* 13 (1989), 229.; Evan Berman, "How Useful is Performance Measurement," *Public Performance and Management Review* 25, no. 4 (2002), 349.; Natalie Buckmaster, "Associations between Outcome Measurement, Accountability and Learning for Non-Profit Organisations," *The International Journal of Public Sector Management* 12, no. 2 (1999), 186-197.

*analyzing the data, and consulting...They have very little time or resources to travel to the field, to understand first-hand the problems, and to assess the quality of data.*⁴⁷

Monitoring is also said to be plagued by perverse incentives or, rather, different incentives for the people collecting data, than the incentives of managers, or leadership, all of which do not necessarily align. The people collecting data may not have the opportunity to use that data and so may not care much about its quality or presentation. There may be few incentives for managers to take action and improve performance if someone above them has more power to make decisions. Moreover, program managers and organization leadership alike may simply feel threatened or reluctant to share findings for fear of revoked funding or lost jobs.⁴⁸ Thus, many scholars argue, monitoring for results and reporting the findings may preoccupy individuals and organizations, resulting in poor quality data, misrepresented findings, or goal displacement where those stakeholders end up more focused on targets than affecting true change.⁴⁹

Relatedly, another oft-cited criticism is the use of over-simplified measurements. Organizations may merely count outputs or even inputs, rather than using indicators to signal outcomes.⁵⁰ In other cases, where change is actually being measured, organizations may place a disproportionate focus on quantitative measures.⁵¹ Also, disparate or ill-fitting data may be aggregated or “rolled-up” into a single report, losing the nuance needed for valid analysis.⁵² Berman hypothesized that organizations may revert to measuring only outputs because measuring outcomes entails stakeholder or customer perspectives which would, in turn, require increased resources and capacity.⁵³ There are likely many reasons for the tendency to measure

⁴⁷ Bamberger, *The Monitoring and Evaluation of Public Sector Programs in Asia: Why are Development Programs Monitored but Not Evaluated*, 229.

⁴⁸ Ibid., 229-231.

⁴⁹ David Bernstein, "Comments on Perrin's "Effective use and Misuse of Performance Measurement"," *American Journal of Evaluation* 20 (1999); Burt Perrin, "Effective use and Misuse of Performance Measurement," *American Journal of Evaluation* 19 (1998), 367.

⁵⁰ David Ammons, "Performance Measurement and Managerial Thinking," *Public Performance & Management Review* 25, no. 4 (2002), 345. Buckmaster, *Associations between Outcome Measurement, Accountability and Learning for Non-Profit Organisations*, 186-197; Perrin, *Effective use and Misuse of Performance Measurement*, 367

⁵¹ Bamberger, *The Monitoring and Evaluation of Public Sector Programs in Asia: Why are Development Programs Monitored but Not Evaluated?*, 223; Berman, *How Useful is Performance Measurement*, 348; Perrin, *Effective use and Misuse of Performance Measurement*, 367

⁵² Ibid.

⁵³ Berman, *How Useful is Performance Measurement*, 349

outputs rather than outcomes, particularly with interventions attempting to catalyze complex changes.

There are other logistical challenges primarily concerning the design of monitoring systems. First, several authors wrote about the unclear definition of terms and concepts that are supposed to be monitored, unclear goals, or unclear relationships between the goal and what is being measured.⁵⁴ The quality of data is a commonly cited challenge regarding monitoring. The data may not answer the right questions, may answer irrelevant questions, or may just be meaningless. Perrin, quoting Sir Josiah Stamp, provides a telling example:

*The Government is very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he pleases.*⁵⁵

These authors do cite some benefits of monitoring as well. The most frequently cited benefit is that monitoring helps improve the functioning of organizations. Some of these have already been discussed as the main purpose for monitoring in the first place, namely to trigger learning cycles, transfer knowledge, and facilitate the formulation of new strategies.⁵⁶ Monitoring may also help determine whether and what type of evaluation is needed.⁵⁷ It can help maintain supporters and convert skeptics through evidence of success which may also help raise funds. Use of monitoring information may help de-politicize resource allocation choices or help defend management decisions by providing objective information.⁵⁸ Monitoring may have a motivational effect by showing implementers the results of their work and may result in increased capacity and ownership.⁵⁹

⁵⁴ Bernstein, *Comments on Perrin's "Effective use and Misuse of Performance Measurement"*; Buckmaster, *Associations between Outcome Measurement, Accountability and Learning for Non-Profit Organisations*, 186-197; Perrin, *Effective use and Misuse of Performance Measurement*, 367

⁵⁵ Ibid., 372; Bamberger, *The Monitoring and Evaluation of Public Sector Programs in Asia: Why are Development Programs Monitored but Not Evaluated?*, 224

⁵⁶ G. Bouckaert and G. Peters, "Performance Measurement and Management: The Achilles' Heel in Administrative Modernization," *Public Performance & Management Review* 25, no. 4 (2002), 360.; Buckmaster, *Associations between Outcome Measurement, Accountability and Learning for Non-Profit Organisations*, 186-197

⁵⁷ Bernstein, *Comments on Perrin's "Effective use and Misuse of Performance Measurement"*

⁵⁸ Ibid.

⁵⁹ Bouckaert and Peters, *Performance Measurement and Management: The Achilles' Heel in Administrative Modernization*, 359

It is important to note that many of the complaints about incentive systems which drive goal displacement, oversimplified measurements, or poor quality data are not flaws of monitoring, as defined in Section IV, but of how monitoring is implemented. These arguments highlight a key finding of this paper which is that oft-cited critiques of monitoring are indeed those of a larger system in which monitoring is situated. The lists of potential challenges or opportunities are long but shallow, reflecting a lack of empirical rigor. Yet, this literature provides a platform from which to begin an analysis of whether or not the factors summarized in the figure below have any bearing in reality.

Challenges	Benefits
Resources: need staff time, capacity building, funds for collection and management, information technology	Increase funds: through budget allocations, fundraising events, support of elected officials
Perverse Incentives: fear of reprisal may lead to manipulating measures, lack of accountability may lead to poor quality data, staff feel threatened by findings	Enhance Support: maintain supporters and convert skeptics through evidence of success
Oversimplified Measures: leads to collection of short-term outputs instead of focus on change, reliance on quantitative measures, or aggregation of data which diminishes nuance	Improve Functioning: improved organizational processes such as enhanced learning, knowledge transfer, strategy formulation
Logistical Challenges: unclear terminology, unspecified methodology, irrelevant or inaccurate measures	Staff Morale and Capacity: increase staff's evaluative capacity and application of critical thinking skills, improve ownership and empowerment

FIGURE 1

V. MONITORING IN PRACTICE

This section explores the experiences and popular sentiments of development practitioners as another body of evidence towards the question of whether and how monitoring contributes value to development programming. This body of evidence is comprised of a combination of three studies on practitioner attitudes, relevant M&E blogs and other online discussions, and informal conversations with practitioners. Commentary will also be provided based on an analysis of the top ten OECD donors' publicly available M&E documents. This will serve as a formal donor perspective in response to many of the concerns voiced by practitioners, although donors' official policies and day-to-day practices may not necessarily align.⁶⁰ The documents were found by a basic search of donor websites and consist primarily of M&E guides, policies, or grant applications.⁶¹

This analysis is constrained by many factors. There is significant variance in the experiences of practitioners, the work they have done, and the agencies with which they have worked. Moreover, the data collected from online discussions, blogs, and from practitioners who agreed to have informal conversations are subject to selection bias. This bias may result from self-selection as those who are active in such discussions reflect certain motivational qualities of the respondents. Also, the fact that the respondents' perspectives were discovered online may reflect something about their location and relative standing in an organization. For example, five of the seven people who participated in informal conversations were working either in North America or the UK and all but one were M&E Specialists or independent evaluators. Their perspective may differ significantly from that of a practitioner working with monitoring on the ground.

In practice, the emphasis on monitoring and reporting during the past two decades is often referred to as the "results agenda." The term "results" in this case means outputs, outcomes, or impact (intended or unintended, positive and/or negative) of a development

⁶⁰ The idea that official policies and day-to-day practices may not align is based on conversations with practitioners who have had to report to donors and on personal observations of publicly accessible quarterly and annual reports to donor agencies.

⁶¹ Synonyms of monitoring such as results-based management and performance measurement were also considered in this analysis, provided they matched the definition presented in the first section. A maximum of two hours were spent exploring each site. If the relevant information could not be found in two hours, it was considered inaccessible and therefore not likely to have significant influence on monitoring, evaluation, reporting, or any related processes.

program.⁶² The term “results agenda” is a subject of heated debate and is related to the similarly contentious debate on “evidence.”⁶³ For this paper, the results agenda refers to the general movement in the international development field for collecting results information in order to demonstrate effectiveness. The Paris Declaration propelled the idea of a results agenda, with more than 100 signatories committing to managing and implementing aid in a way that focuses on the desired results and use of information to improve decision-making.⁶⁴ Three key studies on practitioner attitudes about the results agenda will give important insights to how the previously outlined theories occur in reality.

The most recent study, published in 2013, was conducted by the Big Push Forward website and studied 153 quantitative and 109 qualitative stories of website visitors to better understand their perceptions of the results agenda’s impact on their working lives. While this study has a limited and non-representative sample, it does find quite varied responses about the day-to-day processes and products involving results, what information is being collected, how it is used, and to what effect. The study found overall positive reactions to the results agenda, but points out disparate views on several key issues, such as whether the data collected is useful and what level of detail is necessary for manageable and nuanced data.

In 2012, CDA Collaborative Listening Projects published *Time to Listen*, a book capturing the experiences and voices of over 6,000 people who have received, observed the effects of, or been involved in providing aid. These conversations occurred during the Listening Project which sought to understand the long-term, cumulative effects of different types of international aid efforts and contributed to a growing awareness of needed improvements to aid effectiveness and accountability.⁶⁵

The final study, published in 2009, assesses grant-maker performance through grantee feedback in East Africa.⁶⁶ A questionnaire was used to collect feedback from 305 grantees of eight East African grant-makers on aspects of the grant-makers’ performance such as their

⁶² Marco Segone, *Country-Led Monitoring and Evaluation Systems*, eds. Marie-Helene Adriene and others UNICEF, 56-57.

⁶³ For a synopsis and links to one such debate, see: <http://oxfamblogs.org/fp2p/so-what-do-i-take-away-from-the-great-evidence-debate-final-thoughts-for-now/>

⁶⁴ Organization for Economic Cooperation and Development, "The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action," (2008).

⁶⁵ M. Anderson, D. Brown and I. Jean, *Time to Listen: Hearing People on the Receiving End of International Aid* (Cambridge, MA: CDA Collaborative Learning Projects,[November 2012]).

⁶⁶ Keystone, *Assessing Grantmaker Performance through Grantee Feedback in East Africa*. Keystone, [October 2009]).

funding, application, approval, and reporting processes, non-financial support, reputation and expertise of the grant-maker, and M&E requirements. Additionally, grantees were asked about the quality of the grant-maker's communications and the quality of the relationship in terms of openness of dialogue, mutual learning and trust. For the purposes of this paper, this questionnaire is obviously limited; it focuses only on grantees in East Africa with the majority of grant-makers also being East African organizations. However, it is the only comprehensive study found that gave the perspectives of implementers in the field and their relationship with funders, specifically in the area of monitoring and reporting. Additionally, it is the strongest representation of a non-western perspective used in this analysis.

Monitoring Requires Abundant Resources

As was the case in the theoretical work, concerns about resources used or needed due to monitoring is also a common discussion point among practitioners. Monitoring requires resources to collect, input, and analyze information as well as to establish and maintain a system of use. Human capital is one important resource. Staff need both the time and the capacity to gather and use helpful monitoring information, including collection techniques, storage procedures, and analytic abilities. According to the Big Push Forward study, many practitioners mentioned needing resources such as more staff and also "support from outside consultants, capacity support, training..."⁶⁷ Creating the actual system, particularly if information technology is involved, may become expensive.⁶⁸ These systems may need to be large enough to host data from multiple country offices around the world and may need to be complex enough for people from all of those offices to able to access the data in real time over the internet; however, if resources are constrained, organizations can scale back to what is affordable.⁶⁹ Without a clear system, organizations may be overloaded with data, so they must decide what data are truly needed and worth collecting, while also considering that the need for information may change

⁶⁷ Brendan Whitty, "Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey," *The Big Push Forward* (April 2013), 10.

⁶⁸ Not all monitoring requires information technology, but many organizations do utilize some sort of technology to track data.

⁶⁹ Dave Shellard (Operations & Evaluation Manager, Society for Neuroscience), in discussion with the author, November 2013.

over time.⁷⁰ The expense, with or without information technology, increases with the number of levels at which data must be accessible and the complexity of that data. One can imagine that program officers in a certain region can have relatively easy access to data coming in from their region and can make adjustments as necessary, but as systems become more complex and entire portfolios are monitored, the coordination and analysis of multiple sites, projects, etc. requires a more sophisticated system, thus increasing the cost.

Moreover, the process of designing, implementing, and refining a monitoring system takes time. In its review of results-based management systems in DAC countries, the Working Party estimated that it took five to ten years to fully establish and implement systems which included developing plans, monitoring results long enough to establish trends, and evolving decision-making and reporting processes.⁷¹ Of course, the size, complexity, and technological needs of a monitoring system depend greatly on the type and size of both the program and organization, so it is not possible to state definitively how much monitoring costs.

More frequently, practitioners complain about the abundant use of resources relating to reporting. According to the Time to Listen study,

Aid providers and receivers both say that current report-writing procedures, completely counter to their intent, introduce inefficiencies into international assistance work. Report preparation to meet donor requirements to ensure continued funding for a series of projects becomes an end in itself. Aid agencies say that more and more of their time is spent complying with reporting requirements in each successive year. The costs of reporting have, many feel, overtaken the value of reporting.⁷²

The impact on local partners is also often emphasized, as they may have to deal with the overlapping frameworks, methodologies, data and reporting formats of multiple donors.⁷³ The large-scale review of results-based management in DAC countries used the example of USAID stating, “operating units and implementing partners are beginning to complain that there is no

⁷⁰ John Mayne, "Challenges and Lessons in Implementation Results-Based Management," *Evaluation* 13, no. 1 (2007), 97.

⁷¹ Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*, 134

⁷² Anderson, Brown and Jean, *Time to Listen: Hearing People on the Receiving End of International Aid*, 79

⁷³ Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*, 122

time left for implementing programs, and that much of the higher-order results data collection is not considered directly relevant or useful to them, but is only being used to ‘report upward.’”

However, 80% of the respondents in the East African grant-maker study thought that the demands which grant-makers made for monitoring and reporting their activities and performance were just right. Only 6% of the grantees found the demands to be too much and just around 10% of the grantees specifically mentioned that the funding for M&E had not been adequate. Nearly half of the grantees said their grant-makers provided sufficient funds for them to comply with the monitoring, reporting, and evaluation processes, while 34% said sufficient funds were not provided and 17% received no funds to complete those processes.⁷⁴ While there was no information about the average amount of funding given to these grantees, the field standard for an M&E budget is often considered to be 10% of the overall programming budget. Yet, there is little evidence of this 10% rule in reality.

As for the OECD donors, USAID clearly states in its Evaluation Policy that the agency will devote approximately 3% of total program dollars, on average to external performance and impact evaluation; however, those resources are distinct from those dedicated to monitoring which “may require reallocation away from project implementation, particularly when the opportunity to improve effectiveness through learning is deemed to be very large.”⁷⁵ AusAid also provides a strict number: a maximum of 10% is to be spent on Design, Monitoring, and Evaluation for any project or program and the evaluation cost cannot exceed that percentage.⁷⁶ Although specifics were not included, only one other agency provided any guidance on resources to be spent on M&E activities in general, stating that resources should be reasonable in scope with the utility of information produced.⁷⁷ Again, there may be unofficial policies or norms about the amount spent on M&E, but overall there was little attention paid to application of resources, particularly to monitoring alone. Per discussions with practitioners online, it seems there may be a move away from setting a percentage cap on how much money should be used for a program’s M&E budget because, for example, the funds provided will differ depending on what the program

⁷⁴ Keystone, *Assessing Grantmaker Performance through Grantee Feedback in East Africa*, 29-32 Key to note, however, is that the study does recognize a significant positive bias in the answers to broad questions such as the ones discussed above.

⁷⁵ USAID, *USAID Evaluation Policy* USAID,[January 2011]).

⁷⁶ AusAid NGO Cooperation Program, *Monitoring, Evaluation, and Learning Framework* Australian Government,[2012]).

⁷⁷ Evaluation Department, *Instructions for Evaluation Activities in Norwegian Aid Administration* (Oslo, Norway: The Norwegian Agency for Development Cooperation,[2006]).

seeks to get out of monitoring. The question as to whether there is enough monetary or other support for monitoring remains unanswered; however, as Ebrahim states, “Funders that want nonprofits to measure impacts, but at the same time are unwilling to fund management capacity building and overhead costs for performance measurement, end up undermining both the nonprofits and themselves.”⁷⁸

Conclusion on Resources

Due to the wide variety of implementing organizations' and funder expectations and experiences, there is no conclusive evidence about how much monitoring costs in terms of time, money, and staff capacity, nor is there conclusive evidence that those resources are being properly allocated to allow effective use of monitoring. Donor agencies do not provide much guidance about the amount of resources which should be spent on monitoring or even

M&E more broadly. This could indicate a willingness to be flexible depending on the program, but a lack of strong policy may also mean the activities are being neglected. Finally, implementing organizations and practitioners seem quite concerned about the amount of time and effort which goes to reporting.

Intransigent Targets

This study is primarily concerned with nonprofit organizations that derive their funds either from multilateral aid agencies or from private donors. As was demonstrated in the discussion above, these organizations often employ upward accountability as the dominant form of accountability because their funders require narratives and financial reports to determine whether their assistance has been delivered honestly and without mismanagement, that the target population has been served, and that established goals have been achieved.⁷⁹ The focus on upward accountability seems almost inevitable; as Mary Andersen, et al found,

Aid agencies need resources to do their work, so they appeal to donors. Donors need assurance that aid agencies have well-thought-through plans worthy of their funding. They need to know who will be helped, how, and with what inputs. They

⁷⁸ Ebrahim, *The Many Faces of Nonprofit Accountability*, 24

⁷⁹ M. Anderson, Brown and Jean, *Time to Listen: Hearing People on the Receiving End of International Aid*, 78

need specificity about expected results; they want to know how the agencies will report these results...⁸⁰

Given that this analysis considers monitoring to be more valuable when it benefits accountability of all types, upward, internal, and downward accountability, the current incentive system is a sub-optimal environment for good monitoring. The focus on upward accountability detracts from the potential of monitoring in many ways. First, a common narrative among practitioners is that implementing organizations feel pressure to meet all of the targets set in the design and proposal phases. For example, the Time to Listen report stated, “To receive funding and be seen as “efficient and effective,” agencies accept the limits of their preset plans, often to the detriment of more successful outcomes.”⁸¹ Thus, the study goes on to say, many respondents noted that standardized reporting formats limit honesty and accountability by predetermining reporting categories rather than allowing genuine reflection.⁸²

Individuals within organizations may also experience incentives to ignore or subvert monitoring systems altogether.⁸³ In a review of UNHCR’s results-based management system, some field staff considered the standard indicators of their results framework to be unworkable and simply ignored them.⁸⁴ Eyben paints a more insidious picture of this subversion; she theorizes that program officers or grant managers, whom she calls the ‘squeezed middle’, either mock or vent their anger at the system and then cynically comply.

They try to do the least possible to secure the funding, possibly with a nod and a wink from a sympathetic bureaucrat in the donor organisation who is equally despairing of mandatory requirements. Cynical compliance may sometimes be accompanied by secret resistance: people carry on working according to their own professional judgment, while reporting up the system what they perceive to be ridiculous numbers.⁸⁵

⁸⁰ Ibid. 70

⁸¹ Ibid.

⁸² Ibid.

⁸³ "Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey." The Big Push Forward (April 2013), 15.

⁸⁴ R. Allen and A. Rosi, *Measure for Measure: A Field-Based Snapshot of the Implementation of Results Based Management in UNHCR*UN High Commissioner for Refugees (UNHCR) and Policy Development and Evaluation Service (PDES),[November 2010]).

⁸⁵ Eyben, *Uncovering the Politics of 'Evidence' and 'Results': A Framing Paper for Development Practitioners*, 8

Another anecdote provided in the Big Push Forward survey concurred: "...staff have found ways to 'get around' the new consequences for poorly rated programs [such as] framing a program re-structure as a 'new' program, thus giving the impression that the underperforming program has ended."⁸⁶

These perspectives reflect concerns suggested by the theoretical literature, namely that the focus of implementers becomes reporting successful achievement of targets instead of demonstrating the desired change. Seven of ten donor agencies, however, specifically mention an implementing organization's ability to change, revise, take remedial measures or corrective action, adapt, or adjust. Two of those seven required grant or project applicants to provide a plan for how they would adjust programming if necessary, while the others simply specified adaptation as a use of monitoring information.⁸⁷ For example, DFID's M&E guideline places emphasis on "active learning through application of M&E information to the continuous improvement of strategies, programmes, and other activities."⁸⁸ In addition, AusAid's M&E study states that monitoring "provides the basis of corrective actions, both substantive and operational to improve the program or project design, manner of implementation and quality of results."⁸⁹ Still, DFID is the only agency that gives specific guidance about how adjustments should be dealt with in cooperation with the agency: "If the indicators developed at inception stage are no longer suitable, or if the outputs or activities have changed, the suggested changes to the logframe should be flagged here, [in the annual report narrative], and an updated logframe submitted to DFID for discussions and approval..."⁹⁰

Whether or not funders actually demand that implementers meet all of their predicted targets, if practitioners truly act on this perceived pressure, the value of monitoring is severely diminished. Conducting a needs analysis, careful design of programming and indicators, and running a baseline to establish points of comparison are necessary components of effective programming, and also monitoring. Fabricating expected results may mean that the true effect of

⁸⁶ Whitty, *Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey*, 11

⁸⁷ Ministry of Foreign Affairs, *Model Application Form- Call for Proposals: Human Rights Fund 2014-2017*.

Government of the Netherlands, [2013]).; USAID Learning Lab, *Program Cycle Overview*. USAID, [2011]).

⁸⁸ DFID Central Research Department, "Monitoring and Evaluation: A Guide for DFID-Contracted Research Programmes," *Department for the International Development* (May 2006), 3.

⁸⁹ JR Cook, "Monitoring and Evaluation Capacity-Building Study," *Australian Agency for International Development Sectoral Policy and Review Branch* (1997), 10.

⁹⁰ DFID Central Research Department, *Monitoring and Evaluation: A Guide for DFID-Contracted Research Programmes*, 28

the programming on the target audience is not being properly measured, which has further implications on whether the programming is helping, could be improved, or is causing harm to program participants. The incentive, or perceived incentive, to set and meet inaccurate targets could have negative impacts on the contribution of monitoring, but may have wider implications for development programming in general.⁹¹

Conclusion on Intransigent Targets

Many practitioners believe that to win funding, they must provide and meet specific targets. Moreover, there is a common perception that those targets cannot be adjusted even with changes in programming or context. All but two donor agencies did require implementers to provide indicators or targets in some fashion, but the majority of the agencies also formally state that monitoring information should be used to make adjustments as necessary. Thus, there is some discrepancy between donor requirements and practitioners' perceptions. Perhaps the informal requirements of donors are dissimilar and more influential than official policies. It is also possible that there is a general misperception among practitioners possibly stemming from fear of losing funding or other reprisals.

Oversimplification of Measures

Related to the incentive system outlined above, many practitioners feel that the results agenda, used here as a practical manifestation of monitoring⁹², reduces measurement to assessing lower-order, quantitative outcomes, or to simply counting outputs:

Many funding organizations...are under enormous pressure to deliver attributable results. This emphasis on delivering positive outcomes is important and appropriate,

⁹¹ Even though the humanitarian assistance sector faces its own challenges and is slightly out of the scope of this analysis, a recent study of use of evidence in that sector reveals similar findings: "Situational monitoring and surveillance remains a major weakness in most contexts. However, this study suggests that even if agencies are conducting monitoring, their ability to change the original type of intervention after it has begun is strictly limited. Often, originally proposed interventions are no longer appropriate when it comes time to implement, as so much time passes between implementation and the original assessment. Yet despite knowing that circumstances had changed, they were unable to adapt the program, often due to contractual commitments with donors and hosting governments. This study suggests that there may be little incentive to monitor programs other than as strictly necessary for accountability purposes." J. Darcy et al., "The use of Evidence in Humanitarian Decision Making," *Feinstein International Center ACAPS Operational Learning Paper* (2013), 33.

⁹² While still recognizing that debates surrounding the results agenda encompass much more than monitoring. For example, see <http://blogs.worldbank.org/category/tags/results-agenda>

*but the means of measuring these outcomes must also be suitable. The concern is that the results agenda will in fact drive inappropriate, crude and over-simplified measurement methods with counter-productive effects.*⁹³

There are many ways in which measurements may be oversimplified or reduced to a point that is detrimental to monitoring's purpose. First, the pressure to deliver results, related to the discussion of intransigent targets, means managers may not want to accept responsibility for elements that are out of their control. Objectives are desired changes that are inherently just outside implementers' control, whereas outputs are more easily secured as the products of planned activities. Thus, as Mayne points out, managers have an incentive to set targeted numbers of outputs that they can more assuredly deliver instead of outcomes.

Collecting and analyzing monitoring information at aggregated levels is another way that nuances about a program's progress may be lost. Data that is not disaggregated may result in neglecting an already marginalized population, and the lack of nuance means that the reasons why that population did or did not benefit from the program cannot be known. Often, these are the people in deepest need, as they may be the hardest and most expensive to reach. MacDonald points out that many independent evaluators criticize the use of quantitative aggregation because it hides the richness of changes made and obscures variation which could lead program managers to focus on better-off target groups where results could be most easily achieved. MacDonald indicated, however, that these complaints were not the fault of quantitative measurement but of adopting the wrong measures.⁹⁴ Another reason that organizations might rely more on outputs rather than outcomes could be due to a time lag. It easier to see the immediate products of completed activities, whereas the outcomes may take longer to develop.⁹⁵

The theoretical section pointed out the motivational effect that monitoring can have; sometimes it may be nice for staff members to hear "we have trained 10,000 people this year!"⁹⁶ The tangibility and ease of relying on outputs may be appealing, although it is not useful to assess progress towards desired results. By focusing on a few specific indicators or targets,

⁹³ The big push forward, 2011, <http://bigpushforward.net/archives/1190>.

⁹⁴ Neil MacDonald, "INGO Experiments in Impact Monitoring - DFID Programmes" (Presentation, INTRAC M&E Conference, INTRAC, Gondwana Development Associates, .

⁹⁵ Allen and Rosi, *Measure for Measure: A Field-Based Snapshot of the Implementation of Results Based Management in UNHCR*

⁹⁶ Brian Heilman (Gender and Evaluation Specialist, International Center for Research on Women), in discussion with the author, January 2014.

managers and staff may focus on achieving the numbers and forget what they are really trying to achieve, but it is important to bear in mind that there could be many reasons for this trend to occur.

The evidence provided by donor documents is, again, in stark contrast to concerns expressed by practitioners. Two of the donors published documents that are decidedly output-focused, but six donors required collection of information on changes resulting from programming. The French Agency for Development states that it “has adopted a number of tools and methods designed to enhance the measurement of the results of its operations. They include aggregatable indicators...[which] are numerical indicators that include indicators on achievements and outcomes.”⁹⁷ While this may sound like the numerical indicators could also signal change, the document goes on to explain the focus of measurement to be on tangible products of programming. USAID writes in much clearer language that “the outcome or project “purpose” is the aggregate result of the outputs to be achieved by the project.” These two agencies clearly prefer outputs and aggregate data, but they are outliers. Six of the ten agencies mentioned the need to measure both outputs and outcomes.⁹⁸ CIDA is most clear about why performance should not be measured only by targets:

Targets however... should not appear in outcomes themselves, for a number of reasons. First, when targets are included in a result statement, they limit the ability to report against the achievement of that result by restricting success to an overly narrow window: the target itself. Reporting, in this context, becomes an exercise of justifying why the target was not met or was exceeded (both could be seen as poor planning) instead of comparing expected outcomes to actual outcomes and discussing the variance.⁹⁹

NORAD also addresses this issue directly in their instructions for evaluation activities, stating, “As outputs normally are possible to attribute to the activities performed, there is a tendency for

⁹⁷ French Agency for Development, "Results and Impact Measurement Tools," <http://www.afd.fr/lang/en/home/AFD/developpement-durable/Mesure-par-les-resultats/outils-mesure-impacts> (accessed 1/10, 2014).

⁹⁸ Not all of the agencies used the terms “outputs” and “outcomes”, but the context and definitions provided indicate that the products of activities and changes resulting from the program were the items to be measured.

⁹⁹ Strategic Policy and Performance Branch, "Results-Based Management Tools at CIDA: A how-to Guide," Foreign Affairs, Trade, and Development Canada, <http://www.acdi-cida.gc.ca/acdi-cida/ACDI-CIDA.nsf/eng/NAT-92213444-N2H> (accessed 1/10, 2014).

management to focus on these aspects in the performance monitoring, often neglecting the monitoring of outcomes.”¹⁰⁰

Conclusion on the Oversimplification of Measures

While six out of ten donor agencies mention the need for both outcomes and outputs in design, monitoring, and evaluation phases, two agencies specifically discuss outcomes as outputs and two agencies do not mention measuring either outputs or outcomes at all. Other factors greatly influence the use of outputs or quantifiable information over those which assess change. Managers may be unwilling to take responsibility for change which is inherently out of their control, outcomes may be more difficult to measure than just counting products, and outputs are, therefore, more easily measured in a shorter amount of time. This is detrimental to development programming as people may forget to focus on achieving the changes they seek to create. However, these are not inherent features of monitoring as an activity, but rather of the conditions under which it exists.

Analysis of Accountability and Learning in Practice

It is useful to analyze all of the practical perspectives offered above to understand how and whether they contribute to monitoring’s dual purpose of accountability and learning. The practitioner opinions amalgamated from online discussions, qualitative studies, blog entries, etc. exposed tension in the relationship between implementing organizations and funders, embodied by upward accountability. The Big Push Forward study provides a telling quote: “Such stories confirm the power dynamics at play: that results are at the centre of an accountability system, whereby information in the form of results are packaged up neatly and sent ‘upstream’.”¹⁰¹ This upward accountability, or what Ebrahim calls external approaches to accountability, is important, but has “only limited potential for encouraging organizations and individuals to take internal responsibility for shaping their organizational mission, values, and performance or for promoting ethical behavior.”¹⁰² This paradigm entirely ignores downward accountability.¹⁰³ As Daniel Ticehurst says,

¹⁰⁰ Evaluation Department, *Instructions for Evaluation Activities in Norwegian Aid Administration*, 10

¹⁰¹ Whitty, *Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey*, 15

¹⁰² Ebrahim, *Accountability in Practice: Mechanisms for NGOs*, 816

¹⁰³ The focus on upward accountability also detracts from monitoring being used for internal accountability. There was no explicit attention paid in the literature or current discourse to internal accountability, defined as being horizontal (between different offices or departments). That is likely a symptom of the system which prioritizes reporting upward, rather than sharing within an organization and has implications for organizational learning, as will be discussed later. However, monitoring is defined as an internal activity and most organizational structures inherently establish some system of reporting information to supervisors. For the purposes of this paper, that type of

Good monitoring is pursued by developing communication processes that help understand beneficiaries so as to make aid accountable to them. It does not require rigorous methods geared to academic concerns and obsessive pursuits of attribution of intervention effects. Monitoring has different requirements reaching beyond mechanistic 'feedback loops' to enhance efficiency. It is about developing ways of being accountable to poor people, neglected communities and struggling institutions through listening, gauging how responses vary, feeding back interpretations and taking appropriate action.¹⁰⁴

Downward accountability, a necessary component of good monitoring, has been shown in at least one study to improve the relevance, effectiveness, efficiency, and sustainability of humanitarian and development programs. The joint Christian Aid, Save the Children, and Humanitarian Accountability Partnership report demonstrated that downward accountability mechanisms (sharing information, increasing participation, and opportunities for complaints) increased program effectiveness and built stronger relationships between the participants and implementers. The participants knew more about and understood the purpose behind programming, so they were more likely to buy in, while the communities also trusted the implementers more, felt more respected, and were able to identify and quickly address issues like fraud and mismanagement. Essentially, monitoring has the potential to be a mechanism that projects the voice of participants and provides a system for taking action. However, this function is currently impeded:

Power dynamics influence whose and what knowledge counts and which results matter in development policy and practice. The politics of 'results and evidence' is about how power determines the meanings that influence a course of action, for example what is considered 'robust evidence'.¹⁰⁵

While half of the agencies mentioned stakeholder or participant involvement primarily in the design phase, only one of the agencies truly addressed all of the components of downward

accountability will be considered upward because it is a function of the vertical power dynamic discussed in this section.

¹⁰⁴ Daniel Ticehurst, "The Lost Prophets: A Practitioner's View of M&E," *European Evaluation Society: Evaluation Connections* (September 2013), 11.

¹⁰⁵ Eyben, *Uncovering the Politics of 'Evidence' and 'Results': A Framing Paper for Development Practitioners*, 4

accountability. The Netherlands' Development Cooperation provides a criterion in its model application based on the extent to which partners or the target group can influence a program, how much they are involved in monitoring a program's progress, and what specific influence they have, formally or otherwise, on corrective action.¹⁰⁶ The rarity of this perspective, even in formally stated policy, does not bode well for the actual practices of leading donor agencies. Other agencies such as DFID, NORAD, and CIDA mentioned the importance of allowing ownership and participation by key stakeholders, but this is only one element of downward accountability and leaves out the extremely important element of participants having influence to change the course of programming.¹⁰⁷

The focus on upward accountability dominates true downward accountability and may also mean that monitoring information is used less for learning. As the Big Push Forward study states,

*There is no desire to engage with the results agenda to influence organizational learning. Logic models have been developed recently for programmes but again this is only due to a donor requesting these and they do not feed into programme or organizational learning. The results agenda is only seen as another opportunity to access funding without any appreciation of the wider issues and benefits.*¹⁰⁸

Another practitioner's comment on a blog states, "Local staff began to perceive the success of what they did in terms of their ability to conform to donor reporting requirements rather than listening to and learning from poor people."¹⁰⁹ David Bonbright reported similar findings from a 2006 Keystone survey: 90% of donors felt it was 'essential' or 'important' to have the views of beneficiaries in understanding grantee performance, 83% agreed with the statement that 'more readily available information from the ultimate beneficiaries' perspective would help make better strategic grant decisions,' and 94% felt that it would be 'extremely valuable' or 'valuable' to receive reports from grantees that offer quality feedback from

¹⁰⁶ Ministry of Foreign Affairs, *Model Application Form- Call for Proposals: Human Rights Fund 2014-2017*, 17-18

¹⁰⁷ DFID Central Research Department, *Monitoring and Evaluation: A Guide for DFID-Contracted Research Programmes*, 3;The Norwegian Ministry of Foreign Affairs, *Results Management in Norwegian Development Cooperation: A Practical Guide* (Oslo, Norway: Norwegian Agency for Development Cooperation (NORAD),[2008].; Strategic Policy and Performance Branch, *Results-Based Management Tools at CIDA: A how-to Guide*

¹⁰⁸ Whitty, *Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey*, 11

¹⁰⁹ Commenter response: January 2012, <http://bigpushforward.net/archives/1300>.

beneficiaries about the work of their grantees. However, of those same donors, only 17% routinely require grantees to report on beneficiaries' views on day-to-day performance, only 14% routinely provide grantees with resources to develop the capabilities required to elicit honest beneficiary feedback, and 23% routinely discuss beneficiary feedback with grantees.¹¹⁰

Clearly, there are important discrepancies between what donors express as expectations in official M&E policies and documents and what practitioners perceive to be required of them. In order to analyze the contribution of monitoring, the discrepancy between practitioner perspectives and donor policy needs to be addressed. Are implementing organizations engaging in good monitoring that contributes to learning and accountability? Do they have the power to enact changes based on that information? Below is a matrix of the possible answers to these questions. The answers are simply hypotheses drawn from the practical perspectives discussed above. They do not necessarily represent proven facts.

Practitioner's Perspective: Are organizations engaging in good monitoring?			
Donor's Expectations: Do organizations have the power to change based on monitoring?	Yes		No
	Yes	No	
Do organizations have the power to change based on monitoring?	Yes	Implementing organizations overall engage in good monitoring, using information to learn and report to all stakeholders. Donors support good monitoring and strongly encourage learning and making programmatic adjustments. <i>[Supported by donor documents, but not practitioner perspectives]</i>	Organizations do not have the will, capacity, or resources to engage in good monitoring which contributes both to accountability and learning, but donors believe they should be and try to empower organizations to improve. <i>[No conclusive evidence or ubiquitous support for this hypothesis]</i>
	No	Organizations are frustrated by donor-requested data and reporting, so they subvert the system, giving the donor what they want and otherwise making programmatic changes without permission. <i>[Supported by practitioner stories]</i>	Organizations are, overall, not interested in collecting information to learn and donors are fine with the status quo, desiring to funnel money out without too many programmatic changes. <i>[No conclusive evidence or ubiquitous support for this hypothesis]</i>

FIGURE 2

Figure 2 demonstrates the need to empirically prove the extent to which organizations are engaging in “good monitoring” and donors’ true interest in monitoring activities – both exercises

¹¹⁰ David Bonbright, "Not Learning from Beneficiaries," *Alliance Magazine* 11, no. 2 (2006), 51.

beyond the scope of this thesis; however, it also shows a misalignment of expectations between donors and practitioners. Donors' official policies often state that "good monitoring" is possible and indeed expected, while practitioners feel pressure to subvert the system and do not feel supported in monitoring for downward accountability or for learning. In order for monitoring to contribute the greatest value, these expectations need to be aligned. As it currently stands, implementing organizations generally serve two customers with radically different needs: the participants and the agencies providing resources, both of whom have different expectations of the organization. Oftentimes, the focus of monitoring tends to be on accountability for money received. This may be necessary for an organization's survival, but this donor-implementer relationship creates a power dynamic which tends to affect the way implementing organizations work, set objectives, and whose voice is heard in designing, implementing, or even monitoring and evaluating programs.¹¹¹ Thus, redefining the role of monitoring so that it can contribute to accountability and learning at every level, from participant to donor, will enhance its overall value. As Barder states,

*If you want a focus on results because you think this creates space for local ownership, to enable donors to support the emergence of local solutions and institutions, then we should be thinking about 'post-bureaucratic aid'. Our existing systems have tended to lead to excessive outside prescription and micromanagement; and in principle they should not be needed if we can observe directly the results about which we really care.*¹¹²

Ultimately, there will need to be a wider culture change between both donors and implementers so that monitoring and evaluation practices benefit donors by giving accurate information on the progress of the programs they fund, and benefit the implementing organization by providing accurate and relevant information upon which to make better programmatic decisions. However, the greatest shift will need to be towards accountability to participants, so that the target audience has a say in the design of programs, has recourse for unsatisfactory work, and receives the best possible services. The following section will suggest how the value of monitoring can improve by taking the focus away from strictly upward

¹¹¹ 2013, <http://www.whydev.org/why-your-non-profit-ceo-needs-to-be-the-me-officer/>.

¹¹² Owen abroad, February 2012, <http://www.owen.org/blog/5228>.

accountability and distributing power along the system, enabling organizations to learn and improve.

VI. ANALYSIS OF KEY FACTORS INFLUENCING MONITORING

The framework starting on page 39 was created to analyze the current environment in which monitoring is situated, how that diminishes its potential value, and solutions to enhance its value. The idea is to focus on how monitoring can contribute value to each level along the hierarchy, from participants to field staff to management, leadership, and up to donors. Each actor, detailed in the first column, needs monitoring information for different reasons, but currently only some actors are getting those needs met. The second column, *Towards What*, represents the type of information needed. For example, donors need to know whether the program is achieving its goal and also whether resources are being spent efficiently. Field staff from the implementing organization need to know that information, but they need to know far more detail, down to whether they are achieving key milestones in their work plan. Ideally, field staff would be able to collect only the information necessary for their purposes and then be able to package that data in a way that is useful to other stakeholders.

The third column depicts how the monitoring information from the second column can be used by the corresponding actor to contribute to learning and accountability. However, the current system creates barriers to that potential contribution and these are detailed in the fourth column. The final column gives *Enhancing Factors* which are either actions that the respective level can take or changes that should happen at that level in order to increase the overall contribution of monitoring. In some cases the actors themselves may have the power to enact the suggested enhancing factors, but other potential solutions may be out of that level's control. Enhancing factors primarily include decentralizing decision-making abilities and creating incentive systems for actors at all levels to engage in good monitoring. These two solutions, however, will likely require organizational change, strong leadership support, and integration of monitoring into other organizational processes. Each of these potential solutions will be discussed in further detail in the following section.

A FRAMEWORK FOR ENHANCING THE CONTRIBUTION OF MONITORING				
For Whom?	Towards What?	Potential Contribution	Systemic Barriers	Enhancing Factors
Donor	Progress towards goal, Efficiency	<p>Accountability: how is money being spent, what change is being made, assurance that monitoring of programs is actually occurring</p> <p>Learning: lessons learned for funding portfolio or type of programming; shared lessons with other organizations</p>	The power dynamic created by the funding structure leads to a focus on upward accountability, detracting from downward accountability and learning. Even if donors do not want a sole focus on upward accountability, implementers feel a pressure to please the donor for financial survival. Also, donors need enough information to be accountable to their stakeholders or constituents, but that information needs to be easy to present and compare.	The type of data & reporting requirements need simplified Expectations should be clear through published M&E policies and then they should be enforced Actively and publicly encourage increased focus on downward accountability and learning Fungible budgeting, when appropriate, would allow actors to avoid excessive bureaucracy in program changes
Leadership (Senior Mgmt. or Board of Directors for the Implementing Organization)	Progress towards goal, Efficiency + Progress towards objectives, Context Monitoring	<p>Accountability: how are resources being spent, what change is being made, how involved are participants; leadership provides information up to donors and down to mid-level and field staff about how information is being used and what decisions are being made</p> <p>Learning: are adjustments being made, what type and based on what evidence; leadership should use information to engage in double-loop learning</p>	Systems are not in place for leadership to get the information they need and report it up to donors while also making it useful for staff below them; there are few mechanisms for mid-level or field staff to learn and take action; leadership rarely report down to show how data is being used, what learning is taking place, what decisions have been made	Learning cultures and incentive systems need established so that monitoring is 'mainstreamed.' This will require money, time, and space for integrating DM&E and for learning A system and processes should be established for upward and downward reporting. Leadership should report to staff how the data is used and resulting decisions Require from lower levels only that data which will be used and ideally data that they can utilize as well Implement an easy-to-use data management system and consider creating one person or team responsible for oversight

<p>Mid-level Staff</p> <p>(Management such as country or regional directors for the implementing organization; they may manage several offices and their programs, may be located in country capital)</p>	<p>Progress towards goal, Efficiency, Progress towards objectives, Context Monitoring + Implementation Process: Outputs</p>	<p>Accountability: how are resources being spent, what change is being made, how involved are participants; provides information up and also down to field staff about how information is being used and what decisions are being made</p> <p>Learning: are adjustments being made, what type and based on what evidence; mid-level staff can use information to engage other mid-levels and field staff in double-loop learning</p>	<p>Mid-level managers have been called the squeezed middle, having to report required information up while trying to make use of information in order to improve programming. They may be squeezed by line-item budgets, intransigent targets, or an undercurrent of pressure to show success. Meanwhile, they also do not have the power to change programming, budgeting, or the data collected. This level may be afraid of losing their job or funding for programs if they do not demonstrate positive results.</p>	<p>This level needs increased autonomy to make decisions while still reporting changes upward and downward</p> <p>The system of incentives and learning culture is needed to motivate this level and also so that they can motivate field staff.</p> <p>Role model use of monitoring data, devoting time and money to integrating DM&E and to learning</p> <p>Report to staff how the data is used and resulting decisions</p> <p>Require field staff to collect only necessary data and enforce compliance with organization's data collection system and quality standards</p>
<p>Field Staff</p> <p>(Employees of the implementing organization, actually on the ground, closest to implementation and data collection. May be program coordinators or supervisors.)</p>	<p>Progress towards goal, Efficiency, Progress towards objectives, Context Monitoring, Implementation Process: Outputs and Milestones</p>	<p>Accountability: how are resources being spent; what change is being made; provides information up and also down to participants about how information is being used and what decisions are being made, in direct contact with participants including a mechanism for participants to influence change in programming</p> <p>Learning: taking information, reflecting on it, making small implementation adjustments, consulting with mid-level to make changes to assumptions or the program's theory of change</p>	<p>There may be a lack of resources such as time, money, and possibly capacity to collect and use quality data. The power dynamic created by having to constantly report upwards means these practitioners often feel as though they spend their time reporting data upward while they have no time to actually utilize the data (if it is even useful) and they probably have little power to make changes. They may be or feel responsible for meeting (intransigent) targets and may fear losing their job for not demonstrating positive results.</p>	<p>Field staff should be integrated into designing the program and the M&E plan in order to understand what information they would need to conduct better programming.</p> <p>This level should make implementation decisions and consult upward for fundamental programming decisions that would affect budget by "X" amount.</p> <p>Participate in the learning culture by thinking critically and using data to make informed decisions</p> <p>Ensure field staff have the capacity to collect quality data, input data and utilize the organization's data system</p> <p>Field staff should have the resources and incentives for use of monitoring towards accountability and learning</p>

Participants	<p>Progress towards goal, Efficiency, Progress towards objectives, Context Monitoring, Implementation Process</p> <p>Accountability: participants and field staff should have a mechanism to hold each other accountable where each are provided information about the program, decisions made, and have a mechanism for redress in case either the staff or participants do not uphold their responsibilities.</p> <p>Learning: can contribute explicit and tacit knowledge to program staff and, if involved, help the programming evolve by combining their insights with that of program staff to instigate double-loop learning</p>	<p>It would take a significant amount of time and effort for participants to be involved and informed about every decision made, objective made, activity completed. It is not entirely clear how involved participants want to be. Also, for field staff to give participants information about decisions made or for them to have any credibility in addressing participant concerns, all of the above mechanisms must work. Donors must give info to leaders who give info to mid-level who pass it on to field staff so that they can communicate with participants. Otherwise, the system fails. The power relations in this whole system often prioritize people on top and diminish the value of participants' perspectives.</p>	<p>True weight must be given to participant feedback in all stages of programming and this must be enforced from donors or, at minimum, organization leadership.</p> <p>Accountability mechanisms must be in place for participants to be informed and have some influence over programming choices.</p>
--------------	--	---	--

FIGURE 3¹¹³

¹¹³ Two recent publications used typologies of humanitarian decision making which are similar to this framework. In this framework, the information that leadership utilizes is similar to “Strategic” or “First order” decisions, including organizational strategy, program approach, and funding channel. The information that Mid-Level staff utilize here is similar to “Program Design” and “Planning” decisions, aka “Second” and “Third Order” decisions which include designing the program/intervention, establishing the target audience, and allocating resources. Finally, what ODI classifies as “Operational decisions” are parallel to this framework’s required information for Field Staff and Participants, namely that of program implementation and modification. See: Darcy, J. Humanitarian Diagnostics: The use of Information and Analysis in Crisis Response Decisions: Overseas Development Institute, 2009 and Darcy, J., H. Stobaugh, P. Walker, and D. Maxwell. “The use of Evidence in Humanitarian Decision Making.” Feinstein International Center ACAPS Operational Learning Paper (2013).

Supportive Literature

The following section will explain in detail how decentralized decision-making, incentive systems, organizational culture, and integrated processes can enhance monitoring specifically, but first, a quick review of the evaluation use literature and also some recent studies in the humanitarian sector provides a basis supporting the Enhancing Factors presented in Figure 3.

First, modern program evaluation gained attention earlier than monitoring or performance management, taking shape in the early 1960s when federal funding for and the demand for evaluations of education and antipoverty programs increased.¹¹⁴ Thus, the topic has received more attention from academics, resulting in a greater body of studies and research. In 1986, Cousins and Leithwood's review of 65 evaluations found that quality of methods, type, rigor, credibility of the evaluator or process, the needs of the decision-maker, quality of communication, and timeliness all impacted evaluation use. The study also found that evaluations which focused on a program process or implementation were generally found to be more useful than those dealing strictly with outcomes and that those evaluations which appealed to decision-maker preferences, sought consensus about the evaluation problem, or demonstrated insight into program operations and decisions to be informed had higher rate of use. Finally, Cousins and Leithwood found that use increased when decision-makers perceived a greater need for information and decreased if the results were perceived as a threat to staff.¹¹⁵

Then, Preskill and Caracelli conducted a survey of 282 Evaluation Use Topical Interest Group members of the American Evaluation Association (AEA) in 1997. They found that important strategies for enhancing the use of evaluations were planning for use at the beginning of the evaluation, identifying and prioritizing user needs, considering resource limitations, involving stakeholders, and adequately communicating findings during the process and through reporting.¹¹⁶ Taut and Alkin (2003) conducted a smaller-scale qualitative study of 18 staff members from different hierarchical levels of a university outreach program. The results of these face-to-face semi-structured interviews demonstrated many barriers to evaluation use. In response to an open-ended question about barriers, the participants primarily talked about human

¹¹⁴ Alkin, M., R. Daillak, and B. White. *Using Evaluations*. Beverly Hills, CA: Sage, 1979

¹¹⁵ J. Cousins and K. Leithwood, "Current Empirical Research on Evaluation Utilization," *Review of Educational Research* 56, no. 3 (1986), 331.

¹¹⁶ Preskill, H. and V. Caracelli. "Current and Developing Conceptions of use: Evaluation use TIG Survey Results." *American Journal of Evaluation* 18, no. 3 (1997).

factors, followed by evaluation and contextual factors. Important human factors varied on two themes: evaluator competence and program staff issues. Participants mentioned the importance of an evaluator's social competence such as relationship-building skills, and cited the lack of trust between staff and the evaluator as an important barrier to use. Other evaluation factors which may inhibit use were the lack of access to correct data, inappropriate methods, etc. Finally, important contextual factors included lack of resources, staff turnover, etc.¹¹⁷

In 2006, Fleischer and Christie followed up on the Preskill and Caracelli (1997) survey by conducting a large-scale cross-sectional survey of over 1,000 US AEA members in order to explore evaluators' current attitudes, perceptions, and experiences related to evaluation use. This study found that the top factors influencing use are pre-planning for use, identifying and prioritizing intended users, communicating findings to stakeholders as the evaluation progresses, developing a communication and reporting plan, and interweaving evaluations into organizational processes.¹¹⁸

While there are some important differences between monitoring and evaluation such as the timeline in which each activity is conducted, the person or people who conduct the activity, and the professional standards to which each are held, there are still some transferrable lessons from the evaluation use literature. For example, while staff trust of the evaluator is not exactly an issue in monitoring, the staff's trust in leadership is very important to enhancing the value of monitoring. Other transferrable lessons include carefully planning how information will be used and by whom, applying credible methods with the appropriate level of rigor and accessibility, seeking consensus among users and stakeholders, and integrating monitoring through the organization's staff and other processes. These findings are consistent with the potential solutions depicted in Figure 3, particularly the conclusion that different methods and type of data may be required for different users and that integration into other processes is essential for use.

Recent work in the humanitarian response sector also supports the framework's potential solutions. In particular, the Overseas Development Institute (ODI) and Tuft University's Feinstein International Center have been integral in the discussion of using evidence in humanitarian decision-making. In 2009, ODI prepared a discussion paper for the UN Food and

¹¹⁷ M. Alkin and S. Taut, "Program Staff Perceptions of Barriers to Evaluation Implementation," *American Journal of Evaluation* 24 (2003), 213.

¹¹⁸ C. Christie and D. Fleischer, "Evaluation use: Results from a Survey of U.S. American Evaluation Association Members," *American Journal of Evaluation* 30 (2009), 158.

Agriculture Organization which analyzed how agencies and donors gather and use evidence when determining how to intervene in humanitarian crises.¹¹⁹ The discussion paper has many conclusions on the short-comings of needs assessments which may lead to inappropriate strategies at the start of an intervention, but the conclusions on how decisions are made during the intervention are more relevant to this paper. ODI found that decision-makers needed information to be presented in a concise manner which can be easily understood by those without technical expertise. Moreover, there needs to be more systematic ways to collect and present data so that the amount of raw data is not overwhelming and allows a relatively quick analysis of needs and priorities, in addition to the ability to compare across different sources.¹²⁰

When ODI consulted with decision-makers, donors, UN agencies and NGOs to gain insight into the use of information by decision-makers, they found that information about the crisis had limited relevance and that people's understanding of their organization's decision-making procedures, their implicit values and assumptions, and the mental models with which they processed information mattered as much when actually making decisions.¹²¹ Even when good evidence¹²² was available, ODI found that personal opinion mattered more when making decisions and that untested assumptions about need and vulnerability tended to dominate individuals' and organizations' thinking.¹²³

Then, in 2013, the Feinstein Center published an Assessment Capacities Project (ACAPS) Operational Learning Paper on *The Use of Evidence in Humanitarian Decision Making*.¹²⁴ That study supports many findings from the ODI Humanitarian Diagnostics paper, but provides a deeper analysis into factors influencing specific types of decisions. Previously decided strategic priorities, resource allocation, and existing frameworks and mental models significantly affect decision-making processes by limiting the range of available options. This is

¹¹⁹ J. Darcy, *Humanitarian Diagnostics: The use of Information and Analysis in Crisis Response Decisions* Overseas Development Institute,[2009]).

¹²⁰ Ibid.

¹²¹ Ibid.

¹²² The ODI Humanitarian Diagnostics paper never specifically defines evidence. It provides a typology of three broad categories of information and evidence which are 1) pre-crisis contextual information (capacities, vulnerabilities, livelihood patterns), 2) information concerning the nature of an evolving crisis (early warning, assessment, and monitoring data), and 3) evidence about 'what works' in response to particular kinds of crisis (best practices, standards, protocols). Page 9

¹²³ Ibid.

¹²⁴ Darcy et al., *The use of Evidence in Humanitarian Decision Making* Note: the paper does not provide a very satisfying definition of evidence; it states that information is not in itself evidence but is only 'evidential' to the extent that it supports a given hypothesis or proposition (where it is evidence for something). Page 19

called path dependence and can be seen in programs which roll from year to year without any reassessment of approach.¹²⁵ Moreover, the researchers found that current processes of decision-making tend to be undocumented and transparent, with key assumptions often unstated.¹²⁶ Ultimately, the study concluded that larger, systemic changes in incentives must occur, situational monitoring must be more widely adopted, and the humanitarian system must allow the flexibility for agencies to adapt programs to meet changing needs. While the use of monitoring in the humanitarian response sector, much like that of evaluation use, cannot be entirely translated to monitoring in international development, these bodies of literature can at least show some parallels and give some academic credibility to the conclusions presented in this paper.

With these lessons in mind, the following section will outline in more detail the enhancing factors proposed in Figure 3 and will use three relevant case studies to illuminate those factors. Unfortunately, two out of the three case studies are US-based organizations and are not situated within the international development field; this recurring theme is likely an implication of the overall lack of published resources on monitoring in international development. These case studies were chosen for their level of detail, comprehensive narrative, and the lessons they present, which respond to challenges similar to those that organizations face in international development. First, authority needs to be decentralized so that critically-thinking staff members have the power to actually make programmatic decisions based on sound evidence. Second, an incentive system must be established that encourages organization-wide use of monitoring data for learning and accountability. This, however, will likely require organizational change from a focus on upward accountability toward learning for many institutions. Finally, integrating DM&E processes within an organization will yield the highest value from monitoring.

Decentralized Decision-Making

Both the donor community and implementing organizations need to rearrange their reporting and decision-making structures to be more decentralized and focused on incentives that benefit program participants. The *Meeting Farmer Needs* case on page 47 highlights a scenario

¹²⁵ Darcy, *Humanitarian Diagnostics: The use of Information and Analysis in Crisis Response Decisions*, 11

¹²⁶ Ibid.

where a staff member working in the field observed needed changes, but experienced many of the barriers discussed above, resulting in delayed action.¹²⁷ Ngoma observed that farmers were not willing to participate in the program; it is not clear from the case whether that information was formally stated in any report, or even if it was collected as data. Perhaps Ngoma knew the information by simply being involved in outreach to farmers. Ideally, she would have data on the number of participating farmers and the number who declined, in addition to the reason they declined. This may have provided credibility when it came to proposing an adjustment to the programming. It was not, however, necessarily a lack of codified data that stopped Ngoma from reporting lackluster results. She was afraid of repercussions for admitting failure, but also apparently did not have the authority to make programmatic changes herself. Yet, she was possibly the best-positioned person to make such changes because she had the most access to participants who could tell her the problems and necessary changes to improve the program. It seems plausible, then, that monitoring information is most valuable (in a learning sense) to those closest to the decision to be informed.

Thus, a reasonable solution may be to decentralize decision-making authority to appropriate levels. Different types of decisions could be appropriate for different levels; for example, field staff such as Ngoma may make implementation changes to inputs, activities, and outputs that will change the overall budget by less than a certain percentage. For implementation changes with budget implications, or for changes to proposed program objectives and goals, those staff members need to consult with the mid-level staff member such as the country director or regional manager. The mid-level staff should have the ability make decisions about programs they supervise while still reporting changes upward and downward.

¹²⁷ Charity Ngoma, "Meeting Farmer Needs," Admitting Failure, [http://www.admittingfailure.com/failure/charity-ngoma/](http://www.admittingfailure.com/failure/charity-<u>ngoma</u>/) (accessed 12/29, 2013).

Meeting Farmer Needs

Charity Ngoma, Sector Coordinator for USAID, posted her experience with monitoring on the Admitting Failure website. This example does not explain a formal monitoring system, but demonstrates the value of monitoring which may not necessarily be captured in formal processes. Ngoma managed PROFIT Zambia, a USAID-funded program that sold a vet services package to farmers and veterinarians. In the beginning, the farmers needed to pay a veterinarian for a year's worth of services with the understanding that the vet would come whenever needed. Even though Ngoma noticed that after the first year those who had agreed to pay did not want to pay again, she and her team kept visiting farmers. She reflected on that time,

How do I go and tell my supervisor, or go back to my quarterly report when I am writing to USAID saying, “Oh, we are not getting the numbers, farmers are not buying our interventions.”? It was hard. We didn’t want to do it. So we kept on pushing it. It meant – personally, on me – it meant I was not doing enough. You’re not going in the field enough. You’re not talking to farmers enough. You’re not having enough meetings.

Besides feeling a personal failure, Ngoma was also worried about keeping her job, but the program finally got to the point where they had to admit it was not working. So they started learning by finally asking why farmers were not paying for the services. The farmers did not trust that the vet would actually come and they feared they would lose a significant amount of money. PROFIT then split the original package into smaller packages. Ultimately, it was Ngoma’s supervisor that created an environment for them to learn. Ngoma remembers him saying, “You guys, if it’s not working, come back and tell us it’s not working.” Ngoma writes,

A lot of donor-driven organizations, we have to answer to our donors, of course. Money is coming from somewhere. We’re evaluated based on the numbers: ‘What have you done with the 17 million dollars that have been given to us? But also, this is a lesson to people that are implementing projects. They need to build in their organization a culture of learning. A culture to allow their staff to admit that “No, this is not working.”¹

Ngoma’s story highlights the pressure to report success and demonstrate results which originally led her to avoid reporting accurate results in order to continue programming. This story also clearly shows that sometimes implementers’ perceptions of their supervisors and donor agencies are incorrect; Ngoma’s supervisor wanted the truth and the agency allowed mid-course adjustments to improve the program’s effectiveness. Ngoma was worried about losing the grant or her job, but she started to learn and found success when she finally listened to the data.

A model of decentralized authority, however, assumes that all practitioners are as questioning as Ngoma. It assumes that people all along the hierarchy view their role in the programming cycle with a critical eye, looking for strengths, weaknesses, and opportunities to improve. If that assumption does not hold, decentralized authority will have little to no difference in whether monitoring information is actually utilized. Some practitioners may not have the will to question assumptions and progress. Particularly, for the actors on the lower end of the

hierarchy, they may feel that their job is safer if they simply complete the tasks ahead without criticizing or questioning them.

Thus, if an organization's leadership sees the value in monitoring and desires employees of all levels to utilize monitoring information for accountability and learning, it can establish a new incentive system. This system could involve salary schemes, bonus structures, or other concrete rewards which shows the organization's dedication to good monitoring and makes it worth the time of every employee. For an already existing model, consider DFID's Gender Equality Action Plan (GEAP), published in 2007, which sought to improve DFID's work on gender equality.¹²⁸ The GEAP example shows how technical expertise can be integrated and uplifted by galvanizing leadership support and inducing incentives. The GEAP had many components, including strengthening gender awareness among multinational partners, but the relevant aspects of the plan in this context are its leadership, reporting process, communications, and incentive scheme features. First, gender became the responsibility of senior civil servants working within each division, which gave staff the incentives and encouragement to attend to gender-focused objectives and bring gender into existing work.¹²⁹ Second, the GEAP also included a rigorous reporting process, which is already a feature of most M&E plans, but some regions and countries were able to develop their own relevant and tailored response to the local context. Moreover, there was a consistent initiative to integrate gender-related activities into regional and country plans. Then, DFID made its subsequent two annual reports open to and designed for the public so that outsiders could hold the organization to account for its progress.¹³⁰

Third, DFID carefully implemented communication mechanisms to the entire staff which included a widespread awareness campaign, "Think Women", and a more pointed, practical phase, which included how-to information such as applied examples and technical knowledge-sharing.¹³¹ Finally, the organization implemented a bonus scheme whereby senior staff members were awarded based on their promotion of gender equality. These managers developed their own performance objectives on gender and then wrote a one-page note stating what they had achieved in relation to those objectives at the end of each reporting year. This scheme was controversial,

¹²⁸ Social Development Direct, *Gender Equality Action Plan (GEAP): Light Touch Review Summary Document* UK Department for International Development,[2010].

¹²⁹ Ibid.

¹³⁰ Ibid.

¹³¹ Ibid.

as some thought it was unfair that only senior staff were rewarded and some thought it unethical to pay senior leaders for an issue that DFID should be championing without question. Ultimately, the incentive scheme was designed to create a shift in attitudes and awareness, which it apparently achieved to DFID's satisfaction.¹³²

An incentive system encouraging integration of DM&E must be carefully constructed. This paper has already demonstrated the hazardous effects of linking money to data collection; indeed the donor-implementer system currently embodies this problem by creating a situation where organizations must report certain data in order to get funding (or at least implementers have such a perception). Organizations should not repeat this mistake by prioritizing outputs over change. Therefore, an organization could reward mid-level managers overseeing programs that demonstrated careful design based on a rigorous analysis, with high quality indicators relevant to the desired change, and that had a feasible, yet robust M&E plan. Field level staff could be rewarded on the basis of using rigorous analysis to design programs, setting high quality indicators, and using data to make programmatic decisions.

In a program for prison reform, for example, staff may not necessarily be judged on an entirely changed prison system. They would be judged on the extent to which progress was made toward their program's goal and the effort put forth to collect, analyze, and use data to inform that programming. If a key contextual assumption suddenly failed and the expected change did not occur, the responsible staff member(s) would not necessarily be judged on the failure to create change, but on their response and management in regards to DM&E procedures. Returning to the *Meeting Farmer Needs* case, Ngoma, under an incentive system like the one proposed above, would actually be rewarded for collecting and analyzing data on farmer participation and making the appropriate adjustments to improve the amount of farmers participating. However, Ngoma would also be expected to similarly monitor whether the farmers or veterinarians were experiencing the desired change from the program.

This type of incentive program is not without challenges. First, resources are needed to provide the incentives and also to ensure that each level is adequately fulfilling their monitoring responsibilities and achieving their goals. Also, each actor must have a role in establishing, or at least have a clear understanding of, what is expected of them and then have enough agency to meet those expectations. For example, one proposed idea above was for field level staff to be

¹³² Ibid.

rewarded on the basis of using rigorous analysis to design programs, setting high quality indicators, and using data to make programmatic decisions. In order for the system to work, donors cannot dictate what *rigorous analysis* means or the processes by which staff set indicators and make decisions. Those staff members must agree to the expectations and standards in order to provide legitimacy to the system. Finally, an incentive system should not be instated without the proper support from leadership and/or organizational culture. Without clear leadership, consistent communication, and the correct incentives, linking performance and salaries, budgets, or funding may prohibit preventative programs, inhibit innovation, punish risk taking, etc.¹³³ It is not clear whether decentralization or the proposed incentive system can realistically be implemented, as it requires some restructuring and redistribution of power. However, with the current, vibrant discussion on results and evidence, this may be an opportune time to enact these changes.

Establish a Learning Culture

In order to enhance the value of monitoring for all levels in an organization, the previously suggested decentralization and incentive system must also be supported by a culture of learning. An organization's culture is the set of tacit assumptions shared by a group of people which determines their perceptions, thoughts, feelings, and overt behavior.¹³⁴ To effectively sustain the type of decentralized decision-making and incentive system proposed above, an organization must have a culture that is open to admitting failure, embracing constructive criticism, striving for excellence, and must have strong collaborative processes.¹³⁵ Importantly, systems within the organization should purposefully stimulate and support "the ongoing process of asking questions, the collection and analysis of data, and using what is learned from an inquiry to act on important organizational issues."¹³⁶ While individual implementers are more likely to learn if they have the right information, collective learning requires an organizational culture

¹³³ M. Greenway, M. Hendricks and M. Plantz, "Outcome Measurement: Showing Results in the Nonprofit Sector," *New Directions for Evaluation* 75 (1997), 25.

¹³⁴ H. Preskill and R. Torres, *Evaluative Inquiry for Learning in Organizations*, 1st ed. (California: SAGE Publications, 1998), 155.

¹³⁵ Souvairan, *Why Your Non-Profit CEO Needs to be the M&E Officer*

¹³⁶ L. Botcheva, C. Roller White and L. Huffman, "Learning Culture and Outcomes Measurement Practices in Community Agencies," *American Journal of Evaluation* 23 (2002), 423.

which exhibits routines of data consideration as appropriate organizational behavior.¹³⁷ In building a culture conducive to collective learning, a diverse set of individuals should have the space to inquire, share learnings, uncover assumptions, discuss organizational goals, and present successes and failures through evidence.¹³⁸ The *Assessing Performance at the Robert Wood Johnson Foundation* case below demonstrates the importance of organizational culture in implementing and sustaining a valuable monitoring system.¹³⁹

Assessing Performance at the Robert Wood Johnson Foundation

The Robert Wood Johnson Foundation (RWJF) is the United States' largest philanthropy devoted entirely to public health. As an incoming board member in 1993, Robert Campbell asked, "What are we aiming for, and how will we know when we've gotten there or if we ever will?" Many previous RWJF presidents had pursued attempts at assessing the organization's performance, but President Steve Schroeder (1989-2003) pioneered the organization's drive for continuous monitoring of their grants in addition to the existing individual evaluations.

Schroeder quickly found that he needed to counter the attitude of 'business as usual.' He sought to impart in the staff a restlessness to improve RWJF's performance whenever possible, and cautioned against complacency. Schroeder launched the RWJF Scorecard which originally collected data on the grants awarded, detailed outputs of grants, overall outcome measures, and data on quantitative national measures. However, launching the scorecard was not enough to create an entire monitoring system. RWJF discovered that in order to create institutional change, the leadership had to manage and address staff concerns, many of which were related to issues such as attribution, preoccupation with quantifiable results, and use of resources.

Staff Concerns	Management Response
It is not be possible or appropriate for RWJF to claim credit for impact in the health care field due to the foundation's relatively small and collaborative contributions. Also, a focus on measurement may impede opportunistic grant-making.	The foundation is not implementing a monitoring system to claim credit, but to learn. Having clearer goals and a better understanding of past successes and failures will help to choose the right combination of opportunities.
Most of the foundation's work is not easily quantifiable and to measure it would drive the foundation to less important, but easily measured work.	If direct measurement is too difficult, proxies related to the theory of change will be used and qualitative measures will help elucidate whether goals are being achieved.
Resources spent on the monitoring effort could be better spent on more work, such as more grant-making.	The initiative will take considerable resources, but will ultimately drive learning and improvement, enabling future resources to be better used. Also, resources are needed to build staff capacity in setting goals and explicating theories of change.
The board may not understand the data and improperly use data to make bad decisions.	The goals of staff and the board would be aligned by educating and investing the board in the system from the beginning.

Continued on next page...

¹³⁷ D. Moynihan, "Goal-Based Learning and the Future of Performance Management," *Public Administration Review* 65, no. 2 (2005), 212.

¹³⁸ Moynihan, *Goal-Based Learning and the Future of Performance Management*, 211

¹³⁹ K. Bolduc and P. Giudice, *Assessing Performance at the Robert Wood Johnson Foundation* The Center for Effective Philanthropy, Inc.,[2004]).

RWJF staff was accustomed to using evaluations and had considerable trust in the senior staff's handling of data based on the long history of public grant reporting. Thus, the culture was conducive to addressing this new challenge of monitoring. One of the longest tenured RWJF staff stated,

The staff of RWJF is savvy and knew that the process was real and not going to be used, as it so often is elsewhere, cynically or with ulterior motives. The consistency of senior staff in using the data appropriately, not to undermine people, gave staff the ability to trust that decisions made based on performance assessment data would be constructive, not punitive. That's what has driven the continuation of performance assessment at the foundation.

First, leadership catalyzed organization-wide change that permeated existing structures. The system also had to be introduced strategically and be fully integrated with the foundation's current practices. The staff needed to understand when and how the data would be used and needed to recognize that the monitoring of RWJF's work would clearly inform better decision-making, that it was not purely an intellectual exercise. It was important for management to stress that revealing negative findings was not punitive, but that failures could be constructive and help improve performance or reassess goals. RWJF's monitoring initiative was driven by both the need to learn and to hold themselves accountable. As President Schroeder said, "The inability to measure the bottom line, lack of accountability, unequal power between the foundation and grantees, and people telling you how good you are, leave a tendency to believe that business as usual is the way to go."

It is important to stress that concrete systems and structures need to be created and enforced in order to support staff involvement in dialogue and reflection. Without the support system for staff members, monitoring and evaluation will likely not result in learning.¹⁴⁰ In order to enhance learning, facilitated discussions and other processes can help overcome cognitive biases such as personalization, overgeneralization, or polarization. These biases mean that people tend to relate all people and events to oneself, tend to make generalized conclusions based on one piece of information, or tend to see things as right or wrong, black or white, with no middle ground.¹⁴¹ This was clearly demonstrated in *The Use of Evidence in Humanitarian Decision Making* article published by the Feinstein Center; the mental models, cognitive biases, and preconceived notions of decision-makers mattered as much, if not more, than the actual evidence at hand. Additionally, learning may be constricted by the feeling that time spent on reflection is unduly luxurious, or is not supported by management, or it may be neglected by people feeling comfortable with the status quo.¹⁴² Guided dialogue and reflection, while they may occur in informal spaces, should be required of decision-makers and should be allocated proper time and

¹⁴⁰ Preskill and Torres, *Evaluative Inquiry for Learning in Organizations*, 23-24

¹⁴¹ Ibid., 2

¹⁴² Ibid., 23-24

resources in order to encourage actual learning and use of evidence when designing or making adjustments to programs.

Some concrete learning models that can help transform individual tacit knowledge to explicit collective knowledge were provided in Section IV. In particular, the Knowledge Creation Spiral and Eight Function models provide examples of activities which help organizations learn. These include establishing times for staff to interact through communities of practice, creating internal guidelines or procedures from research or evaluation findings, and initiating communication systems for shared learning. Facilitated discussions can also help encourage staff or other actors to ask questions in order to acquire information, insight, clarity, and direction that helps resolve problems more efficiently and effectively while capturing learning.¹⁴³ The RWJF leadership encouraged its staff to ask these sorts of questions throughout the implementation of the system and consistently valued staff questions and opinions. This contributed to an already strong evaluative culture and is related to the idea of decentralization; staff members will feel more empowered if their individual learning is valued, if they are encouraged to ask questions, and if their opinions and questions receive response from leadership. Preskill and Torres write, “The consistent and ongoing questioning about the practices, processes and outcomes of our work stimulates continuous learning, a sense of connectedness, and improved individual, team, and organizational performance.”¹⁴⁴

Many practitioners believe, and management literature generally agrees, that one of the most important factors in establishing an organization’s culture is strong leadership. Whether an organization’s orientation leans more towards accountability or towards learning depends on external influences and the organization’s preexisting history and culture, but leadership also has a role to play. A leader that focuses on evidence-based decision-making may encourage risk-taking, experimentation, and learning, while a focus on accountability and reporting may encourage more conservative behavior.¹⁴⁵

The importance of leadership is clearly demonstrated in the RWJF case where the Board and President catalyzed the implementation of a new monitoring system and worked to build a culture that supported it. To create more of a balance between accountability and learning, an

¹⁴³ Ibid., 2

¹⁴⁴ Ibid., 2

¹⁴⁵ Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*, 121

organization's leadership will need to influence the culture from the top down by creating intentional opportunities to learn and improve. Leadership can accomplish this by setting an example of using data to make decisions and also by providing staff the time, resources, and open environment to share results, good or bad. Measurement must be seen as a structural necessity.¹⁴⁶ The leadership can also set a vision that helps members know what to focus on, where to put resources, and how to choose or reject certain opportunities.¹⁴⁷ Preskill and Torres write, "It is important that employees routinely hear and see their leaders engaging in learning activities, talking with others about learning, and planning future learning initiatives."¹⁴⁸ However, if an organization does not have leadership which champions monitoring and evaluation, coalitions of supporters may nonetheless be able to change aspects of the organization's culture. Internal networkers or community builders are often less visible in these types of organizations, but they may seek out others who are open to developing new learning capabilities.¹⁴⁹

Integrating Monitoring: Through Process and People

In order for changes such as decentralizing authority and altering incentive systems to successfully improve the value of monitoring, implementing organizations must also increase their capacity and willingness to practice good monitoring. M&E must be seen not only as a technical department, but as a necessary part of an organization's work.¹⁵⁰ As Bamberger states,

*Most project M/E systems are required to produce information for a large number of government and donor agencies, and often the M/E unit will not even be consulted as to the kinds of data to be collected or how it is to be used. Thus, in some cases the M/E unit is reduced to little more than a data collection service. The kinds and volume of information requested will often bear no relationship to the staff and financial resources available for data collection.*¹⁵¹

¹⁴⁶ J. Eckhart-Queenan and M. Forti, *Measurement as Learning: What Nonprofit CEOs, Board Members, and Philanthropists Need to Know to Keep Improving* The Bridgespan Group, [April 2011].

¹⁴⁷ Preskill and Torres, *Evaluative Inquiry for Learning in Organizations*, 2

¹⁴⁸ Ibid., 2

¹⁴⁹ Ibid., 2

¹⁵⁰ Erica Stillo (Senior Monitoring and Evaluation Specialist, Plan), in discussion with the author, November 2013.

¹⁵¹ Bamberger, *The Monitoring and Evaluation of Public Sector Programs in Asia: Why are Development Programs Monitored but Not Evaluated?*, 229

Organizations should work to integrate monitoring within the entire program cycle, starting with the design of a program, through evaluation, and into the organization's learning processes.¹⁵² The Bridgespan Group has analyzed several case studies on how to integrate monitoring into programs for enhanced effectiveness and concludes that four dimensions are necessary: define, measure, learn, and improve. When decision-makers define the results they want to achieve and how to get there, they are better able to figure out what information they need to measure results, they can learn what works and what does not, and then explicitly apply what they have learned to improve their results.¹⁵³ The case, *Our Piece of the Pie: From Data to Decision-Making*, demonstrates these important steps.¹⁵⁴

Our Piece of the Pie: From Data to Decision-Making

The Bridgespan Group published a case study on *Our Piece of the Pie* (OPP), a nonprofit organization based in Hartford, Connecticut. In 2005, their biggest contract, a federal grant, was ending, so the management team commenced a strategic planning process which resulted in re-focusing on being a youth-serving organization with one core program called "Pathways to Success" and one additional youth-focused employment program. The organization had the capability to collect individual beneficiary data, but there were many issues such as inaccurate data, difficulty identifying mistakes, and problems aggregating the data. Moreover, OPP's financial and operational data were stored in several places, the data collection methods varied by department, and the staff completed timesheets which were not coordinated with anything else. This all led to frustrations as people could not determine if they were making the best use of their resources, could not quantify the cost of each type of service, and could not predict the resources they would need.

Thus, the Bridgespan Group, hired to unlock the value of data in managing OPP's programming, began by asking OPP management what decisions it would inform with better data, and which pieces of data were required to support these decisions; they also needed to know the most effective and efficient process to supply data in a timely manner. Everyone suffered from a data deficit; the youth development services did not know how well youth adhered to their services, if the services even aligned with youth goals, or if the Pathways program was succeeding because of its design or for other reasons. The education coordinator wanted to know if staff members were allocated optimally and the Vice President of Administration wanted to track the costs associated with Pathways to analyze whether certain groups were more expensive to serve than others, and how changes in the composition of youth may alter financial needs over time.

OPP management identified what data they would need to inform decisions such as program performance and resource allocation, then they determined a different set of metrics for each decision to be made, and finally set up an Excel system that could host all of the necessary kinds of data. The system was made to be user-friendly for the staff to collect, analyze, and report data. OPP created audience-specific "dashboards", adjusted how they collected and reported some data, and created the Quality Assurance Director position, responsible and accountable for inputting data, identifying data gaps or mistakes, and providing dashboard reports to the management team.

Continued on next page...

¹⁵² Dave Shellard (Operations & Evaluation Manager, Society for Neuroscience) and Luc Bourgie (Evaluation Consultant) in discussion with the author, November 2013.

¹⁵³ Eckhart-Queenan and Forti, *Measurement as Learning: What Nonprofit CEOs, Board Members, and Philanthropists Need to Know to Keep Improving*, 7

¹⁵⁴ A. Cortez and L. Vetter, *Our Piece of the Pie: From Data to Decision-Making* The Bridgespan Group,[January 2008].

After four months of using dashboards regularly in management meetings, the management was able to foster staff compliance and proficiency with the new system, errors were more obvious and easier to be identified, investigated, and corrected. The system helped senior management make better use of its time with front line staff, and helped the staff to more clearly see the connection between the accuracy of data and their ability to serve clients. Specific decisions were also made as a result of the monitoring system. For example, the management observed that the youth development specialists were spending more time with those youth attending school than those youth who were out of school without a diploma; this was opposite of the desired intervention, so they opened a dialogue with the specialists and decided to reallocate caseloads to balance high- and low-needs youth. Also, the ability to combine timesheets, accounting data, and the number of youth in each program allowed management to find that two services varied widely in costs so management could reallocate resources to the most cost-efficient program and serve more youth with the same budget.

OPP learned that a dashboard is only as good as its data, so resources needed to be invested to improve data integration and data-entry compliance. This took time and was challenging for some staff members. Establishing the Quality Assurance Director position was crucial to oversee data entry and to create dashboards, but it took six to twelve months for staff to take ownership of the new system and another six to twelve months to gain true proficiency. The Director reflected on the entire experience:

We are strengthening our capacity and building a data-driven learning organization...A simple flip of the switch solution is not possible. Only a journey focused on clear-headed results with relentless open-minded attention to detail and the human beings driving results will work.

The OPP case reinforces three key lessons from the evaluation use literature which are that intended use must be planned in advance, all stakeholders should be involved in designing the data plan, and an integrated system throughout the organization is needed to enhance use. Monitoring should be linked to design because it is easier to collect useful information with a clearly defined goal and theory of change and when indicators are logically linked to desired results. Monitoring should also be integrated with evaluation practices because both activities often share the dual purpose of learning and accountability, both usually require clearly defined goals and a theory of change, and both practices can share at least some of the same indicators and data. Integration of monitoring into other organizational processes such as design and evaluation gives it context and relevance for staff at all levels. Design, monitoring and evaluation can each add value to a program, but they are far more valuable when integrated into one system. One holistic system is both dependent upon and can enhance staff and management buy-in.

Specifically, monitoring and evaluation have many complementarities which can be exploited to save resources and enhance staff buy-in for evidence-based decision-making. Rist outlines most of the commonly known complementarities which include sequential, information,

and organizational complementarities.¹⁵⁵ Monitoring information may generate questions to be answered by evaluation studies, and is often useful in summative evaluations to gain a sense of what actually occurred during programming (sequential).¹⁵⁶ Also, both monitoring and evaluation may draw from the same data sources and use that information for different audiences or analyses (information). Finally, monitoring and evaluation information may be channeled through a centralized system rather than through distinct units which could mean that evaluators may contribute to the development of a monitoring system or may help improve indicators and/or data-collection methods (organizational).¹⁵⁷ Nielsen and Ejler add methodological complementarity to Rist's framework which means both monitoring and evaluation share similar processes and tools for structuring and planning, obtaining data, analyzing and inferring judgment.¹⁵⁸ Those authors argue that quantitative data are easier to be used with both M&E, while evaluations may be supplemented with more in-depth techniques such as interviews and participant observation.¹⁵⁹

While evaluation will likely need more data points to triangulate evidence and may indeed utilize more in-depth techniques, monitoring practices may also include techniques such as participant observation which yields qualitative instead of only quantitative data. At the heart of integrating design, monitoring and evaluation is the idea that each step simultaneously depends on and enhances the others. These complementarities show how DM&E build off of each other and how that benefits decision-makers in their planning and aligning of resources to achieve intended results.¹⁶⁰ Moreover, integrating DM&E means that all staff, from field staff to mid-level staff and leadership (and participants to the extent possible) must have an understanding of the relationship between a program's goal, objectives, and activities. It is also necessary to include these actors in indicators, collecting, analyzing and reviewing data,

¹⁵⁵ R. C. Rist, "The 'E' in Monitoring and Evaluation: Using Evaluative Knowledge to Support a Results-Based Management System," in *From Studies to Streams: Managing Evaluative Systems*, ed. Rist, C and Stame, N. (London: Transaction Publishers, 2006), 3-22.

¹⁵⁶ Hatry, *Sorting the Relationships among Performance Measurement, Program Evaluation, and Performance Management*, 24

¹⁵⁷ Nielsen and Ejler, *Improving Performance? Exploring the Complementarities between Evaluation and Performance Management*, 174

¹⁵⁸ Ibid., 174

¹⁵⁹ Ibid., 176-177

¹⁶⁰ Lahey and Nielsen, *Rethinking the Relationship among Monitoring, Evaluation, and Results-Based Management: Observations from Canada.*, 52

conducting evaluations, and using information to make decisions.¹⁶¹ If the measurement tools are logically linked with the program, of high quality, and deemed appropriate by participants, staff, and management, the relevant actors will be more likely to see the system as credible and relevant. Thus, they will be more willing to comply with it.

People's perceptions of the value of monitoring likely vary by the role they play, and the type of experiences they have with monitoring.¹⁶² This is demonstrated in the OPP case study where staff of all levels clearly understood the project logic and also what specific data they needed to make decisions. Actually using the new system and being held accountable for compliance took time, effort, and increased capacity, but that level of integration across processes and staff was vital to increasing the value of monitoring information. Some people may resist the idea of monitoring their work as they could feel that years of training and experience cannot be replaced by a system that reduces each aid project to indicators and targets. They may also feel that using these forms of measurement as incentives or accountability mechanisms unjustly questions their intentions.¹⁶³ Thus, an organization needs to involve people at all levels of the organization in engaging with the data and creating the expectations and energy for data use because "when groups of people in an organization have an intimate knowledge of the data and have argued about its meaning and applicability, they have a possibility of developing a shared purpose and working together to reach their goals."¹⁶⁴

The RWJF and OPP case studies also demonstrated that this level of integration necessitates increased technical capacity across all staff members, so supporting mechanisms must also be put in place. These may include training exercises, establishing central units responsible for providing technical assistance, building databases with consistent, well-organized, timely, and easily accessible monitoring information, developing conceptual frameworks, and creating procedures for planning and reporting.¹⁶⁵ For staff at all levels, data should be understood as a necessity in deciding a course of action; however, it is also necessary to give individuals, particularly those motivated to collect and utilize monitoring information, the

¹⁶¹ Binnendijk, *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*, 127

¹⁶² "Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey." The Big Push Forward (April 2013), 6.

¹⁶³ Owen abroad, April 2012, <http://www.owen.org/blog/5483>.

¹⁶⁴ Stephanie Sutherland, "Creating a Culture of Data use for Continuous Improvement: A Case Study of an Edison Project School," *American Journal of Evaluation* 25 (2004), 280.

¹⁶⁵ Ibid.

power to act on the things they learn.¹⁶⁶ Indeed, Preskill and Torres write, “By underutilizing employees, we create cultures of apathy, lethargy, and anti-learning. The truth is, we have failed to tap the potential of organization members’ knowledge and experience.”¹⁶⁷

Summary of Enhancing Factors

The graphic below serves as a summary tool for the Enhancing Factors discussed above. The framework in Figure 3 analyzed the current and potential contributions of monitoring and provided some factors that could enhance the value of monitoring. This section has discussed the factors in detail, but the main points are synthesized in Figure 4. Each factor – decentralized decision-making, establishing a learning culture, and integrating monitoring – can help enhance monitoring in its own right, which is represented by the colored squares. However, each of the enhancing factors also supports one another.

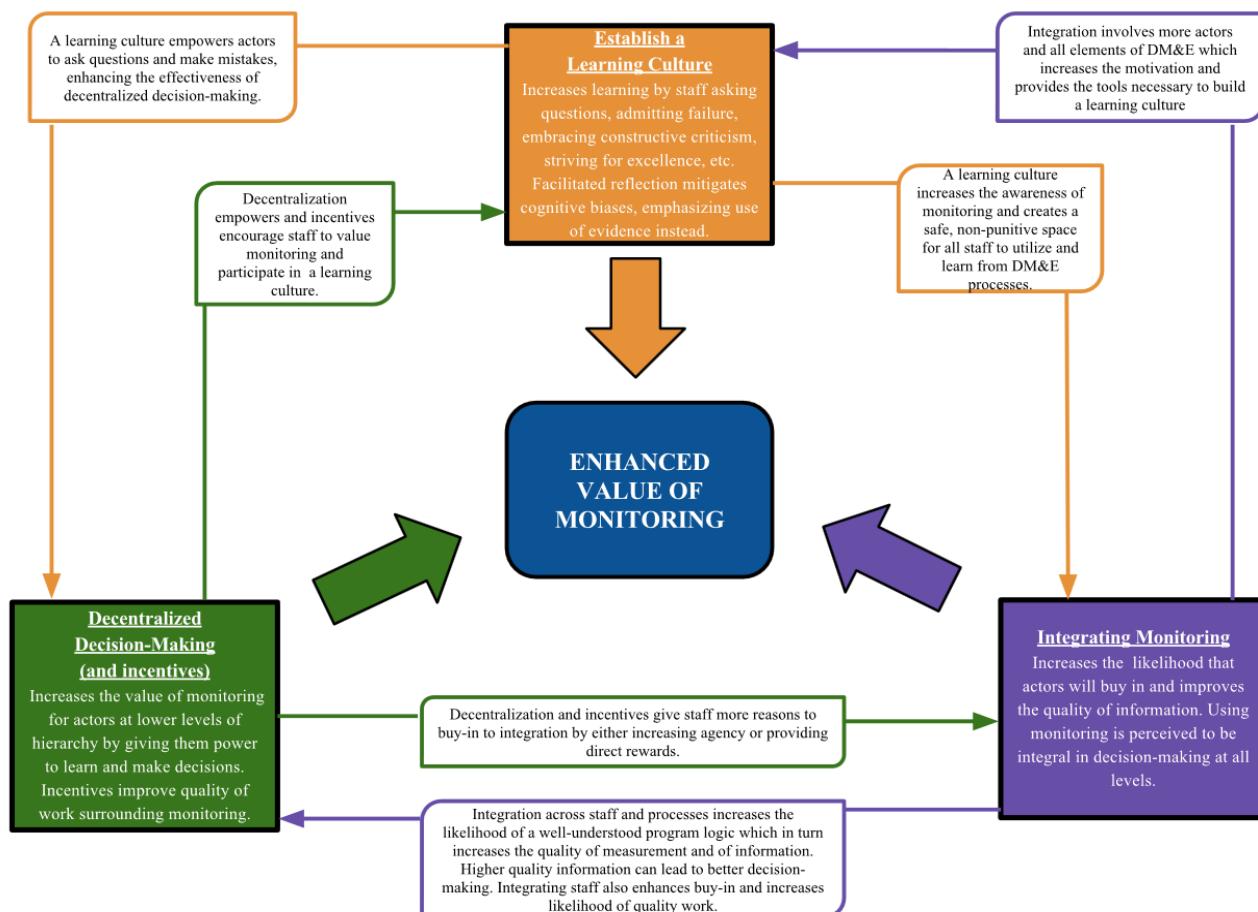


FIGURE 4

¹⁶⁶ Moynihan, *Goal-Based Learning and the Future of Performance Management*, 212

¹⁶⁷ Preskill and Torres, *Evaluative Inquiry for Learning in Organizations*, 63

VII. CONCLUSION

There is a scarcity of empirical evidence demonstrating whether monitoring has contributed value to international development programs. Moreover, the relevant theoretical work on monitoring and performance measurement lacks rigor and fails to adequately address challenges or benefits of monitoring. In fact, the available academic literature and professional discourse provide little evidence that “good monitoring” is being conducted in the field. Good monitoring occurs when the intended use of monitoring information and the ways in which it will be collected and analyzed are planned during the program design phase. Additionally, information should be collected on project implementation, progress toward results, and contextual relevance in order to hold stakeholders accountable, and to facilitate learning aimed at improving programming. There is a rich debate occurring around issues of monitoring, but upon closer examination, much of it is in fact a conflation of the ‘results-agenda’ and general reporting mechanisms, rather than a critical assessment of the benefits and challenges of monitoring itself. The common critiques of the current donor-driven results-agenda, listed below, include the amount of resources used for data collection and report writing, intransigent targets, and an oversimplification of measures.

- Monitoring requires money, time, and staff capacity to collect, input, analyze, store, and use information which may or may not include information technology as an added expense. While many practitioners may feel burdened by reporting requirements to the point that no time remains for implementation, or for collecting data on higher-order results, there is no definitive evidence about how much time or money is or should be spent on monitoring activities. This information is highly dependent on the type of program and the type of information needed for accountability and learning purposes. A lack of clear budget guidance from donor agencies may indicate flexibility in the available resources for monitoring activities, or may mean those activities are, in fact, neglected.
- Many practitioners perceive such intense pressure to meet all the targets set in the proposal phase that they feel unable to collect information and process it for their own reflection. This may lead staff members to subvert the current monitoring system by either reporting inaccurate information or simply ignoring indicators they feel are irrelevant to their work.

The majority of donor policies directly contradict these practitioner perceptions, emphasizing the importance of using monitoring information to learn and adapt programming. Thus, either informal donor requirements are dissimilar from and more influential than official policies, or there is a misperception among practitioners.

- The same perverse incentive system described above also leads to oversimplified measurements, either by counting outputs instead of measuring changes, by focusing only on quantitative indicators, or by aggregating important and nuanced information. The donor policies are more mixed regarding this issue, with just under half of the agencies specifically mentioning the importance of measuring objectives and goals, while two other agencies themselves conflate the idea of outputs and outcomes. Managers and implementing organizations may also have many reasons for counting outputs instead of measuring change; the implication of oversimplifying measurements, however, is that monitoring loses significant value in terms of learning.

Even though these criticisms have little to do with the flaws of monitoring, as this paper defines it, the barriers presented within the system nonetheless diminish its potential value by ignoring the fact that monitoring should provide more than just upward accountability. Currently, monitoring is rarely used for downward accountability or learning, and at the heart of that failure are the power dynamics between donors, implementers, and participants. The analysis showed significant discrepancy between donor and practitioner attitudes and expectations. For monitoring to contribute its potential value to both learning and accountability, attention needs to be reallocated so that implementing staff and participants perceive to have and actually acquire more power to make decisions. This will require a shift in power dynamics, but it can ultimately benefit all actors, as donors will have more accurate information on the programs they fund, implementing organizations will have relevant information upon which to make better informed decisions, and participants will have the knowledge and power to influence those decisions.

Thus, both implementing organizations and donors can take some concrete actions to enhance the value of monitoring. Donors should create consistent demand and requirements for good monitoring, while helping to shift the focus towards downward accountability and learning. They can also initiate a system of decentralized decision-making as illustrated in Figure 3.

Implementing organizations also have a responsibility to reestablish regulations, expectations, roles, and responsibilities, in order to give more autonomy to the staff member(s) closest to the decisions to be informed.

The idea of autonomous decision-making relies on the assumption that the majority of practitioners are willing and able to critically analyze the program cycle with the intention of learning and improving. However, if that is not an accurate analysis of most practitioners or organizations, an incentive system may be established which links internal rewards to staff members who demonstrate progress towards personal objectives and use of evidence for decision-making. Both decentralization and a new incentive system would require increased staff capacity and continual support systems to build and maintain M&E skills and to foster a culture of inquiry and learning. A supportive learning culture should include opportunities for admitting failure, giving and embracing constructive criticism, in addition to strong collaborative processes during planning and design phases. All of these should purposefully stimulate and support asking questions, analyzing data, and actually using what is learned to improve organizational or programmatic functions. This should certainly entail deliberately facilitated trainings, dialogues, and reflections, but may also include informal opportunities to collaborate with colleagues. An organization's leadership must actively devote the communication space, resources and enthusiasm for monitoring, in addition to modeling good practice by clearly establishing the organization's goals, choosing quality indicators to measure progress, implementation quality, and assumptions, and making decisions based on high-quality data.

Finally, implementing organizations should integrate monitoring into their full program cycles. This involves establishing a strong program design with clear goals and a theory of change. Then, a monitoring and evaluation plan should be jointly created, including generating indicators based on the theory of change, establishing data collection methods, and clarifying roles and systems for the collection, analysis, interpretation, and sharing of information. Those systems need to be routinized and accepted by leadership and a critical mass of supportive staff. Organizations may also consider creating a position such as a data quality manager who would be in charge of checking data quality, but could also facilitate the creation of a monitoring system that caters both to reporting and learning. All of the "Enhancing Factors" – decentralized decision-making and incentive systems, establishing a learning culture, and integrating

monitoring – can individually enhance the value of monitoring, but they are also all interconnected and support each other.

The three case studies collectively demonstrate how these recommendations can improve programming by restructuring the services offered, better serving target audiences, and utilizing resources more efficiently. These examples show how monitoring can in fact positively contribute to international development. However, thus far, there is no concrete evidence that the majority of development organizations are engaging in “good monitoring”, and, there is no proof that they are benefitting from their current monitoring practices. In order to better understand and potentially increase the value of monitoring on a larger scale, further research would be useful; some important research questions may be:

- How many organizations are engaging in good monitoring? What types of organizations are engaging in good monitoring and do they represent certain sectors or types of programming? Are certain sectors more inclined to continually monitor programs and their context and adjust as needed (e.g. highly dynamic programs in conflict settings)?
- To what extent are individual practitioners engaging in good monitoring which is not detected by formal mechanisms? Are these individuals stifled by reporting, subverting the conventional reporting systems, or working within the systems?
- To what extent does funding diversity affect implementing organizations’ monitoring practice? Does increased diversity correlate to practices that are necessary for good monitoring, in particular monitoring that is used for learning in addition to accountability?
- What types of funders or which specific funders best facilitate “good monitoring” practice? What do funders actually expect or desire in terms of monitoring information? Are they averse to organizations changing the implementation, theory of change, and/or underlying assumptions of funded programs? If so, why? What would this mean for the use of monitoring?
- How many resources are organizations truly putting in to monitoring activities (or M&E more broadly)? Does this differ for organizations engaging in “good monitoring” versus organizations that may only count outputs or collect information only for donor use? What are the actual benefits of monitoring to all stakeholders (donors, implementers, participants, etc.)? Is it possible to conduct a cost-benefit analysis for monitoring and compare that across organizations that have different conceptions of monitoring?

These research questions could be explored either through large-scale surveys of donors, implementers, and participants across multiple sectors or types of programming. This would of course require variation in represented donors, geographic location, and size of implementing organization, among other things. These questions could also, to some extent, be explored

through in-depth case studies which follow the path of data collected or decisions made from implementation up the entire system. With more robust evidence, the field can make a more informed decision about whether the monitoring and evaluation industry is indeed worth less than a pitcher of spit, as Owen Barder so eloquently stated. This analysis has shown, however, that with the right conditions in place – by giving more power to lower levels of the donor-implementer hierarchy, providing incentives to effectively utilize monitoring, establishing a supportive learning culture, and integrating monitoring into organizational processes – monitoring can have a positive influence on both learning and accountability.

ANNEX A: ACKNOWLEDGMENTS

- I. A series of informal conversations with practitioners were conducted in November 2013 over Skype and the practitioner names are printed here with their permission. All of these individuals contacted the author in response to a request posted on the Pelican Initiative, a Platform for Evidence-based Learning & Communication for Social Change. Even though each individual was not necessarily quoted or referenced, each helped tremendously in helping formulate the thinking and analysis behind this paper.
 - **David Shellard**, Operations & Evaluation Manager at the Society for Neuroscience
 - **Erica Stillo**, Senior Monitoring and Evaluation Specialist at Plan
 - **Guy Sharrock**, Senior Advisor for Monitoring & Evaluation at Catholic Relief Services
 - **Luc Bourgie**, Consultant
 - **Madelene Barboza**, Consultant at Instituto Fonte
 - **Mansoor Ali**, International Project Manager at Practical Action
 - **Rhonda Schlangen**, Independent Evaluation Consultant
- II. Additionally, the author greatly appreciates the help and insights of several Fletcher alumni including Brian Heilman, Elise Garvey, Elizabeth Wright, Margot Isman, and Sophia Dawkins.
- III. Last, but not least, significant gratitude must be extended to Professor Cheyanne Scharbatke-Church for providing guidance, encouragement, and insights to this paper. Without her Design, Monitoring, and Evaluation (DM&E) series at Fletcher, none of this would have happened!

BIBLIOGRAPHY

- Alkin, M., R. Daillak, and B. White. *Using Evaluations*. Beverly Hills, CA: Sage, 1979
- Alkin, M. and S. Taut. "Program Staff Perceptions of Barriers to Evaluation Implementation." *American Journal of Evaluation* 24, (2003): 213.
- Allen, R. and A. Rosi. *Measure for Measure: A Field-Based Snapshot of the Implementation of Results Based Management in UNHCR*: UN High Commissioner for Refugees (UNHCR) and Policy Development and Evaluation Service (PDES), November 2010.
- Ammons, David. "Performance Measurement and Managerial Thinking." *Public Performance & Management Review* 25, no. 4 (2002): 344.
- Anderson, M., D. Brown, and I. Jean. *Time to Listen: Hearing People on the Receiving End of International Aid*. Cambridge, MA: CDA Collaborative Learning Projects, November 2012.
- AusAid NGO Cooperation Program. *Monitoring, Evaluation, and Learning Framework*: Australian Government, 2012.
- Bamberger, Michael. "The Monitoring and Evaluation of Public Sector Programs in Asia: Why are Development Programs Monitored but Not Evaluated?" *Evaluation Review* 13, (1989): 223.
- Barber, Owen. *The Lethal Effects of Development Advocacy*. Owen Abroad. November 2009
- Barber, Owen. *Seven Worries about Focusing on Results, and how to Manage them*. Owen Abroad. April 2012.
- Barber, Owen *What are the Results Agenda?*. Owen Abroad. February 2012.
- Berman, Evan. "How Useful is Performance Measurement." *Public Performance and Management Review* 25, no. 4 (2002): 348.
- Bernstein, David. "Comments on Perrin's "Effective use and Misuse of Performance Measurement"." *American Journal of Evaluation* 20, (1999).
- Binnendijk, Annette. *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*. France: DAC Working Party on Aid Evaluation, 2000.
- Bolduc, K. and P. Giudice. *Assessing Performance at the Robert Wood Johnson Foundation*: The Center for Effective Philanthropy, Inc., 2004.
- Bonbright, David. "Not Learning from Beneficiaries." *Alliance Magazine* 11, no. 2 (2006): 51.

Botcheva, L., C. Roller White, and L. Huffman. "Learning Culture and Outcomes Measurement Practices in Community Agencies." *American Journal of Evaluation* 23, (2002): 421.

Bouckaert, G. and G. Peters. "Performance Measurement and Management: The Achilles' Heel in Administrative Modernization." *Public Performance & Management Review* 25, no. 4 (2002): 359.

Britton, Bruce. "Learning for Change: Principles and Practices of Learning Organizations." *Swedish Mission Council* no. 5 (2002).

Britton, Bruce. "Organisational Learning in NGOs: Creating the Motive, Means and Opportunity." *The International NGO Training and Research Center Praxis Paper No. 3*, (March 2005).

Buckmaster, Natalie. "Associations between Outcome Measurement, Accountability and Learning for Non-Profit Organisations." *The International Journal of Public Sector Management* 12, no. 2 (1999): 186-197.

Carman, Joanne. "Nonprofits, Funders, and Evaluation: Accountability in Action." *The American Review of Public Administration* 39, (2009): 374.

Causton, A., C. Hagens, D. Morel, and C. Way. "Guidance on Monitoring and Evaluation." *Catholic Relief Services* (October 2012).

Christie, C. and D. Fleischer. "Evaluation use: Results from a Survey of U.S. American Evaluation Association Members." *American Journal of Evaluation* 30, (2009): 158.

Church, C. and M. Rogers. *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*.
Search for Common Ground, March 2006.

Cook, JR. "Monitoring and Evaluation Capacity-Building Study." *Australian Agency for International Development Sectoral Policy and Review Branch* (1997).

Cortez, A. and L. Vetter. *Our Piece of the Pie: From Data to Decision-Making*: The Bridgespan Group, January 2008.

Cousins, J. and K. Leithwood. "Current Empirical Research on Evaluation Utilization." *Review of Educational Research* 56, no. 3 (1986): 331.

Darcy, J. *Humanitarian Diagnostics: The use of Information and Analysis in Crisis Response Decisions*: Overseas Development Institute, 2009.

Darcy, J., H. Stobaugh, P. Walker, and D. Maxwell. "The use of Evidence in Humanitarian Decision Making." *Feinstein International Center ACAPS Operational Learning Paper* (2013).

DFID Central Research Department. "Monitoring and Evaluation: A Guide for DFID-Contracted Research Programmes." *Department for the International Development* (May 2006).

Donaldson, Stewart. "A Practitioner's Guide for Gathering Credible Evidence in the Evidence-Based Global Society." Chap. Epilogue, In What Counts as Credible Evidence in Applied Research and Evaluation Practice?, edited by Donaldson, S., C. Christie and M. Mark, 239. California: Sage Publications, Inc., 2009.

Ebrahim, Alnoor. "Accountability in Practice: Mechanisms for NGOs." *World Development* 31, no. 5 (2003): 813.

Ebrahim, Alnoor. "The Many Faces of Nonprofit Accountability." *Working Paper, Harvard Business School* 10-069, (2010): 1.

Eckhart-Queenan, J. and M. Forti. *Measurement as Learning: What Nonprofit CEOs, Board Members, and Philanthropists Need to Know to Keep Improving*: The Bridgespan Group, April 2011.

Evaluation Department. *Instructions for Evaluation Activities in Norwegian Aid Administration*. Oslo, Norway: The Norwegian Agency for Development Cooperation, 2006.

"Experiences of the Results Agenda: Draft Findings for Discussion from the Crowd-Sourcing Survey." *The Big Push Forward* (April 2013).

Eyben, Rosalind. "Uncovering the Politics of 'Evidence' and 'Results': A Framing Paper for Development Practitioners." *The Big Push Forward* (April 2013).

French Agency for Development. "Results and Impact Measurement Tools." , accessed 1/10, 2014, <http://www.afd.fr/lang/en/home/AFD/developpement-durable/Mesure-par-les-resultats/outils-mesure-impacts>.

Greenway, M., M. Hendricks, and M. Plantz. "Outcome Measurement: Showing Results in the Nonprofit Sector." *New Directions for Evaluation* 75, (1997): 15.

Hatry, H. P. "Sorting the Relationships among Performance Measurement, Program Evaluation, and Performance Management." *New Directions for Evaluation* 137, (2013): 19.

Herrero, Sonia. *Integrated Monitoring: A Practical Manual for Organisations that Want to Achieve Results*: inProgress UG, April 2012.

Keystone. *Assessing Grantmaker Performance through Grantee Feedback in East Africa*: Keystone, October 2009.

Kusek, J. and R. Rist. "Ten Steps to a Results-Based Monitoring and Evaluation System." *The International Bank for Reconstruction and Development/The World Bank* (2004).

Lahey, R. and S. Nielsen. "Rethinking the Relationship among Monitoring, Evaluation, and Results-Based Management: Observations from Canada." *New Directions for Evaluation* 137, (2013): 45.

Lee, J. "NGO Accountability: Rights and Responsibilities." *CASIN: Programme on NGOs and Civil Society* (2004).

LeRoux, K. and N. Wright. "Does Performance Measurement Improve Strategic Decision Making? Findings from a National Survey of Nonprofit Social Service Agencies." *Nonprofit and Voluntary Sector Quarterly* 39, no. 4 (2010): 571.

MacDonald, Neil. "INGO Experiments in Impact Monitoring - DFID Programmes." Presentation, INTRAC, Gondwana Development Associates.

Mayne, John. "Challenges and Lessons in Implementation Results-Based Management." *Evaluation* 13, no. 1 (2007): 87.

Ministry of Foreign Affairs. *Model Application Form- Call for Proposals: Human Rights Fund 2014-2017*: Government of the Netherlands, 2013.

Moynihan, D. "Goal-Based Learnng and the Future of Performance Management." *Public Administration Review* 65, no. 2 (2005): 203.

Murtaza, Niaz. "Putting Lasts First: The Case for Community-Focused and Peer-Managed NGO Accountability Mechanisms." *International Society for Third-Sector Research* 23, (2011): 109.

Ngoma, Charity. "Meeting Farmer Needs." Admitting Failure, accessed 12/29, 2013, <http://www.admittingfailure.com/failure/charity-ngoma/>.

Nielsen, S. and N. Ejler. "Improving Performance? Exploring the Complementarities between Evaluation and Performance Management." *Evaluation* 14, (2008): 171.

Organization for Economic Cooperation and Development. "The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action." (2008).

Pasteur, Katherine. *Learning for Development: A Literature Review*. No. 6: Brighton: Institute of Development Studies, 2004.

Perrin, Burt. "Effective use and Misuse of Performance Measurement." *American Journal of Evaluation* 19, (1998): 367.

Preskill, H. and V. Caracelli. "Current and Developing Conceptions of use: Evaluation use TIG Survey Results." *American Journal of Evaluation* 18, no. 3 (1997): 209.

Preskill, H. and R. Torres. *Evaluative Inquiry for Learning in Organizations*. 1st ed. California: SAGE Publications, 1998.

Rist, R. C. "The 'E' in Monitoring and Evaluation: Using Evaluative Knowledge to Support a Results-Based Management System." In *From Studies to Streams: Managing Evaluative Systems*, edited by Rist, C and Stame, N., 3-22. London: Transaction Publishers, 2006.

Roche, Chris. *Reporting and Aggregating Results* The Big Push Forward, January 2012.

Segone, Marco. *Country-Led Monitoring and Evaluation Systems*, edited by Marie-Helene Adriene, Petteri Baer, Michael Bamberger, Osvaldo Feinstein, Enrico Giovannini, Denis Jobin, Megan Grace Kennedy, et al UNICEF,.

Social Development Direct. *Gender Equality Action Plan (GEAP): Light Touch Review Summary Document*: UK Department for International Development, 2010.

Souvairan, Emanuel. *Why Your Non-Profit CEO Needs to be the M&E Officer* Why Dev, 2013.

Strategic Policy and Performance Branch. "Results-Based Management Tools at CIDA: A how-to Guide." Foreign Affairs, Trade, and Development Canada, accessed 1/10, 2014, <http://www.acdi-cida.gc.ca/acdi-cida/ACDI-CIDA.nsf/eng/NAT-92213444-N2H>.

Sutherland, Stephanie. "Creating a Culture of Data use for Continuous Improvement: A Case Study of an Edison Project School." *American Journal of Evaluation* 25, (2004): 277.

The Norwegian Ministry of Foreign Affairs. *Results Management in Norwegian Development Cooperation: A Practical Guide*. Oslo, Norway: Norwegian Agency for Development Cooperation (NORAD), 2008.

Ticehurst, Daniel. "The Lost Prophets: A Practitioner's View of M&E." *European Evaluation Society: Evaluation Connections* (September 2013): 10.

USAID. *USAID Evaluation Policy*: USAID, January 2011.

USAID Learning Lab. *Program Cycle Overview*: USAID, 2011.

Whitty, Brendan. *Driving Change from within*. The Big Push Forward. 2011.

Winston, Jerome. "Performance Indicators - Promises Unmet: A Response to Perrin." *American Journal of Evaluation* 20, no. 1 (1999): 95.