

Shaping International Evaluation

A 30-Year Journey

Edited by
Gary Anderson



Montreal and Ottawa
CANADA

UNIVERSALIA MANAGEMENT GROUP
5252 de Maisonneuve Blvd. W., Suite 310
Montreal, Quebec H4A 3S5
CANADA
www.universalia.com

Copyright © 2010 by Universalia Management Group. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means – electronic, mechanical, photocopying or otherwise – without the prior permission of Universalia Management Group.

Canadian Cataloguing in Publication Data

Main entry under title:

Shaping International Evaluation: A 30-Year Journey

ISBN 0-9739343-2-8

Copy editor: Tracy Wallis

Cover photo: Marie-Hélène Adrien

Printed in Canada

This publication is printed on recycled paper.

CONTENTS

SECTION I

Introduction

Chapter 1: Shaping International Evaluation	3
---	---

SECTION II

Cross-Sectoral Trends Influencing International Evaluation 1980-2010

Chapter 2: Something Called Capacity: The Evolution of an Idea in International Development	19
Chapter 3: Results-Based Management	45
Chapter 4: Evaluating Partnerships in the Not-for-Profit Sector	75
Chapter 5: The Evolution of Institutional and Organizational Assessment	113

SECTION III

Evaluation Trends in Specific Sectors and Regions

Chapter 6: Evaluating Agricultural Systems	159
Chapter 7: Changing Perceptions of the Environment and Environmental Evaluation	205
Chapter 8: Evaluating ICT: A New Dress for Old Questions	239

Chapter 9: Evaluating Gender Equality: Universalia Experiences and Perspectives	265
Chapter 10: How M&E is 'Perceived' and 'Applied' in the MENA Region	295

SECTION IV

Conclusion

Chapter 11: Conclusion	317
Contributors	329
Index	335

SECTION I

Introduction



CHAPTER ONE

Shaping International Evaluation

*Gary Anderson, Marie-Hélène Adrien,
and Charles Lusthaus*

Universalialia began in 1980, in an era when Microsoft and Apple were less than five years old. There were no cell phones or fax machines, and the worldwide web had not yet been conceived. It was a time when secretaries typed evaluation reports, though computers were beginning to be introduced for word processing as well as their traditional role in statistical calculation. Universalialia consisted of Charles Lusthaus and Gary Anderson, the two founders, Gerry Cooney, our first professional employee, and a receptionist / secretary / bookkeeper. Few organizations of any kind can boast that the first three paid professionals are still engaged with the organization 30 years later – perhaps not even Microsoft and Apple, though they have strengths that Universalialia has yet to attain!

Evaluation has evolved considerably since Universalialia was founded 30 years ago. Then, evaluation was just beginning to organize as a discipline. The Canadian Evaluation Society was not formed until 1981, and the American Evaluation Association came into being only in 1986. Universalialia is proud to have been part of the development of the evaluation field in general, and the international development

evaluation field more particularly. To celebrate our 30th Anniversary, we invited a number of influential development writers and evaluators to join our celebration by contributing a chapter to this book. These contributions have allowed us to celebrate our 30th anniversary in the Universalia spirit of “theory and practice”, and to share our thoughts along with our exceedingly busy colleagues. We feel very honoured and privileged to have their original contributions and we thank them all!

Introduction to the Book

The book has two main sections. The first deals with major trends that have influenced international evaluation, and begins with Peter Morgan’s thoughts on capacity development, a dominant theme in development agencies, but as Peter points out, a hollow theme in international agreements, and one whose fuzziness creates great issues for the evaluator. Ken Stephenson follows with his analysis of Results-Based Management, another major influence on international evaluation. Born in the same timeframe as Universalia was created, RBM supported the rising tide of conservatism espoused at the time by Margaret Thatcher in the UK and Ronald Reagan in the USA. However, like the context in which it arose, RBM has limitations in practice. The following chapter deals with another strong international development influence, “partnerships”. Charles Lusthaus, Katherine Garven, and Silvia Grandi have recently conducted substantive work in this area, so Chapter 4 enables them to share what they have learned. Finally, in response to the limitations of a “project” mentality, and in recognition of the importance of developing organizations and institutions, in Chapter 5, Katrina Rojas and Charles

Lusthaus consider how such structures are evaluated. This is a lasting contribution of Universalialia's work over more than two decades, and provides a logical conclusion to the first main section.

The second substantive section considers the evolution of evaluation within specific sectors. We admit that the choice of sectors was largely arbitrary, but has worked out well. From the days of the Colombo Plan, agriculture has been seen as key to international development, so we begin the second section with Ronald Mackay and Doug Horton's chapter on evaluation related to agricultural systems. It is perhaps fair to say that these authors have long been ahead of thinking in the field they evaluate, so they are ideal candidates to share their insights. Nowadays, environment is an important theme in development circles, and we are pleased to have Ramon Pérez Gil share his insights on the evolution of environment as a dominant international force as well as trends in evaluating environmental impact. Two additional "sectors" are included, Information and Communication Technology (ICT) and Gender Mainstreaming. Ricardo Gomez and Shaun Pather look at ICT, a new, dynamic, and exceedingly important theme in our modern world, while Katherine Garven, Katrina Rojas, Anette Wenderoth, and Elisabetta Micaro consider trends in gender mainstreaming and the evaluation response. We conclude the second section with Doha Abdelhamid's insights on development in general, and particularly developing evaluation capacities in the Middle East and North Africa. That contribution nicely melds the concepts of capacity, organizational development, and the dominant themes of the book.

Reflections on What is Ahead

In considering the theme of capacity development, Peter Morgan introduces the notion that development agencies latch on to fads and adopt a passionate pursuit of such slogans, which in themselves are abstract and very difficult to define. While definitions are challenging enough for native speakers of English, understanding the meaning of capacity development may be an insurmountable problem for speakers of other languages. This, of course, introduces even more challenges for the evaluator. If people aren't sure what it is, and cannot define it, it is next to impossible for the evaluator to ascertain whether an organization has it or not! Furthermore, if "capacity" is really "capability", then it represents a potential for performance rather than current performance. Thus, the evaluator rapidly crosses into the sphere of predictive science – a domain fraught with risks and uncertainties.

Morgan also points out the demands of increasingly complex development problems such as climate change for systems thinking and solutions. The linear forms of log-frames and cause-and-effect thinking simply do not fit with the need to understand complex, loosely coupled webs of interactions within equally complex enabling environments.

As well, there may be a huge gulf between Western donor views of capacity and what people in developing countries value and understand. The enhancement of community engagement through elder's circles or increasing the reach of traditional medicine healers may be of great importance to villagers, but neither are understood nor valued by development agencies. The indigenous systems of rewards and cultural patterns governing how those systems work are

essential. Evaluators seldom consider these informal systems.

No matter how we approach capacity development, evaluation has yet to accommodate the hard to measure human attributes that are the real determinants of success or failure – such attributes as drive, will, persistence, leadership, adaptability, and inspiration. Educators often marvel at the success of a student who on the things measured by school examinations was at best average, but who goes on in life with determination and passion that enables him or her to become a major leader and contributor to great enterprise. Trying to emulate science, evaluators sometimes try to control “extraneous” variables, or ignore them in their analyses; however, when these are the primary contributors to success or failure, such an approach is folly. In organizational performance terms, it is what Universalia has termed “motivation” in the organizational performance framework. That has always been a significant contributor to performance, and it has always been and remains the most difficult aspect to assess.

In international development, the disappointment of many programs to achieve their intended results led to much soul-searching and the application of approaches that had promoted success in the early days of the Space Program. These frameworks strengthened the logic and in favourable circumstances enabled program evaluation to increase understanding of how programs functioned including their impact on beneficiaries. Alas, as Ken Stephenson explains in the chapter on RBM, this new orthodoxy failed to live up to its proponent’s expectations. The logic models that promised both learning and accountability were usurped by “false RBM”, which for a wide range of organizational and economic reasons actually increased compliance and

centralized control rather than encouraging program iteration based on evaluation findings.

We have worked with RBM, trained people in its application, and evaluated many projects in terms of their underlying logic. In some cases, projects have flawed logic. The results chain may have illogical cause-and-effect relationships, the assumptions may be unrealistic, and so forth, in which case a project evaluated that is based on flawed logic, is doomed from the start – at least as far as achieving its planned objectives. Hopefully, by forcing people to think through what they intend to do, and to analyze pitfalls and risks, RBM can introduce some common sense in development. One must be cautious of any model that remains static or imposes overly precise or silly indicators to measure and monitor useless information. While RBM may have begun as a useful tool to help people think about development, in too many instances today, it has become a technocratic tool. As Stephenson points out, development agencies began to apply RBM to their central function with results chains that aggregated their programs and projects. This may be good in theory, but aggregating projects and programs sometimes makes little logical sense.

Thirty years ago, the world was a simpler place since many of the presenting problems were local and bounded. As we know today, many of the world's greatest problems cut across sectors, boundaries, and interest groups. They are not easily addressed by single entities working in isolation. As time has passed, there is an increasing requirement for organizations to join forces, through such mechanisms as networks, joint ventures, and consortia. Lusthaus, Garven, and Grandi have authored a chapter on evaluating such groupings of organizations in the not-for-profit sector, which they refer to generically as “partnerships.” While partner-

ships take many forms, they tend to be goal-driven, hierarchically flat entities formed to serve the interests of the partners. The chapter examines the various stages of development of partnerships as well as a number of factors required for success. Because of their increasing importance in international development, evaluation of partnerships is crucial. However, evaluation cannot rely simply on more of the same evaluation approaches used to evaluate projects and programs. Evaluating effectiveness is complicated by multiple goals. In the first place, we need to examine whether the partnership is achieving its intended goals, but we also need to consider sustainability of the partnership itself as a goal, for often it is the partnership that becomes the sustainable way of addressing development issues. Evaluating efficiency is also complicated, especially because partnerships often have very high transaction costs. Other challenges in evaluating partnerships include issues of when in the development of a partnership it should be evaluated, how partners can best participate in evaluation processes, and how to formulate a basis of judgment for success. This is a new area for evaluators, but Lusthaus, Garven, and Grandi provide helpful insights on which evaluators can build in the future.

Development agencies have long recognized that fragile contexts require institutions and organizations as well as projects, programs and partnerships. Just as it became important to differentiate between effective and ineffective programs, it became important to differentiate between good and not-so-good organizations. Over two decades ago, senior evaluators at Universalia realized that development funding was incorporating core support to organizations who acted as key development partners for development agencies. For example, CIDA was contributing project funding to Canadian NGOs at that time, and many of them

lacked fundamental capacities such as robust financial management and knowledge of M&E.

CIDA wanted to provide core funding to these organizations and asked Universalialia to develop a guide on evaluation of NGOs that would be used to judge which NGOs would be eligible for core funding. The guide became a standard tool in CIDA. Shortly thereafter, IDRC began collaborating with Universalialia to understand the effects of its funding of research organizations around the world. Whereas CIDA's interest was in external evaluations, IDRC's interest was in self-assessments. Rojas, Garven, and Micaro portray the full account in the chapter on organizational assessment.

Unlike the RBM approach, which based M&E on quantitative indicators, the IDRC/ Universalialia OA approach focuses on understanding organizational performance using both qualitative and quantitative measures. Thus, it uses a wide variety of mixed methods to build understanding of what is recognized as a dynamic, loosely coupled system in an ever-changing environment.

Over the years, we have found that the organizational assessment framework was also relevant to partnerships since partnerships typically involve two or more organizations. These are new "organizational forms".

One advantage of having conducted evaluations for over 30 years is the opportunity to learn particular approaches to evaluation research in-depth and move with trends that are in fashion in a particular era. We started our early evaluation work when quantitative research methodologies were seen as the "gold standard". Gary Anderson, as a member of the Evaluation Group of Harvard Project Physics, a National Science school curriculum being developed in the 1960s, was able to apply sophisticated quantitative methods to evalua-

tion of the new physics curriculum being developed at Harvard at the time. The aim of the program was to design a humanistically oriented physics course in order to make an inherently interesting subject relevant to students and to attract more students to the subject. Agricultural research and the statistical approaches of R.A. Fisher were the standard in educational evaluation at the time, and while it was possible to explore certain innovative questions, the best questions were deemed those that lent themselves to robust statistical methods. The evaluation team did not think of sitting with young adolescents and asking them about the perceived lack of appeal of the subject, its probable irrelevance to many students and competing alternatives; however, in the end the evaluation of the project provided insight to the researchers to improve their products and programs. The evaluation also taught the group that parents, schools, children, and ministries had questions about education and learning that even the most advanced statistical methods were ill equipped to address.

It is refreshing to learn, as described by Mackay and Horton in their chapter on agricultural research, that education was not the only field swept up in the euphoria of what seemed to be “pure” science using empirical analysis of easy-to-measure variables. Reading their analysis has enlightened us about a field that we are not experts in, but it provides a familiar parallel to the evolution of evaluation in education, a discipline we have worked within, on and off, throughout our careers. They tell us that quantitative research designs can answer some questions but far from all.

MacKay and Horton point out that what counts traditionally is what can be measured well and what can be easily controlled. Mackay and Horton’s anecdote of Dr. Ceccarelli’s conversation with a peasant farmer who chose the opposite

types of plants to the Biological Scientist struck a strong cord. Mackay and Horton conclude that the evaluation questions come first, and that certain methods are better suited to certain questions.

Evaluation is contextual, subject to the trends and fads of the thinking of the day. Nowhere is this more apparent than in the environmental sphere, for that sector has changed dramatically over the past 40 years. Indeed, the environment has moved from just another sector to a multi-dimensional force, which is one of the major drivers in our contemporary world. Pérez Gil traces the evolution of environment and how people have interacted with their natural world. He chronicles the rise of organizations and institutions that shepherd the environment, and proposes various paradigms that have underpinned thinking about the environment, and consequently how we evaluate environmental questions. In early days, people were not concerned with environmental degradation, for they believed that it was resilient and able to absorb myriad abuse. Evaluation in that era was an afterthought, and occupied little space. As the world gained knowledge and understanding, the environment became more salient to diverse interest groups. Indeed, it eventually became a battleground, pitting the resource industries against those bent on preservation. Governments responded by legislating standards for end-of-the-pipe discharge, and the evaluation mantra was one of monitoring and compliance evaluation. Of course, such simplistic paradigms were an inadequate response to complex systemic problems. Thus, as the world began to understand the complexities, evaluation had to adopt more sophisticated approaches that matched the ecology of the world with equally interwoven ecological evaluation methods.

Universalialia developed the evaluation framework for the Montreal protocol on ozone depleting substances during this era. It followed a logical linear results chain in which UNEP funded technological change that removed destructive production processes from the world and replaced them with more environmentally friendly technologies. One could count the tons of emissions reduced, and could derive a global impact on ozone-depleting substances. In the end, that initiative was a significant success, and helped the world correct the mistakes that had produced a dire set of conditions. The major challenge was to ensure valid reporting and that destructive technologies were permanently removed from industries. This was effective evaluation within that context, though limited in terms of overall understanding in the environment sector at large.

Universalialia subsequently supported such organizations as IUCN and the GEF in their more holistic strategies for environmental rehabilitation. We built evaluation capacities to help these organizations understand the effectiveness of what they were doing, and we supported them in conveying an evaluation message to their constituents. In particular, we promulgated the use of organizational assessment to support the performance of organizations throughout the world, which were actually doing the work on the front lines. The environment sector begs some of the most complex questions of our day, and so far, the evaluation community has not done a thorough job of articulating them in useful nested sequences.

A completely different set of issues surrounds the evaluation of ICT programming. As Gomez and Pather point out, ICT is a means to development, not development per se. We like their analogy with electricity, which when installed in a developing country can be used for myriad purposes, for

people's betterment, or for their oppression. Indeed, the infrastructure itself is moot on how it will be used, so elaborate logic models and forecasts of economic benefits are merely dreams. Specified indicators and economic forecasts may bear little relationship to reality. More infrastructure and more access say nothing about how ICT will be used. They point out the cycle that characterizes ICT at differing levels of maturity, from a facility, service, or partner to an enabler of development. Even the efforts of leading development agencies have not fully cracked the conceptual problems of evaluating this field. The authors suggest that in future we should explore frameworks, which focus less on direct and tangible economic, social, or political benefits and more on indirect, intangible benefits. These are more difficult to measure and point to deeper and more pervasive effects on individuals and communities. Thus, much more work needs to be done, opening up opportunities for exploration in future decades.

Reading Gomez's and Pather's chapter on ICT reminded us of a major evaluation *Universalia* conducted a few years ago for Canada's Department of Foreign Affairs, which examined Global Peace and Security in several troubled spots in the world. The aim of creating peace in contexts that have already descended to or recently emerged from war is difficult to comprehend. For example, providing helicopters for African Union troops in Darfur is certainly well intentioned, but it is by no means convincing that their presence contributes to peace. As "observers", the soldiers were not allowed to intervene when innocent people were being assaulted, raped, or killed by armed rebels. The presence of these foreign troops also had highly negative environmental effects, including depletion of water sources, especially when the troops were found to be consuming ten times the water that local people consumed, not to speak of barring

water access for livestock. Not only is “peace” difficult to assess, but in this field, more often than not, it was “absence of war” that was the desired program effect. How does one evaluate the effect of a program on preventing violence?

Such questions abound in development evaluation. We recently evaluated the education system of UNRWA, the UN organization responsible for Palestine refugees in five host countries and territories. That was an interesting evaluation on many levels, but the values issues stand out. What is an acceptable standard of education for a UN organization educating refugees? Is it sufficient to reach the level of host countries when that level is low in world terms? Given a chronic shortage of funds, can a UN organization compromise on health and safety standards for pupils in the schools it manages? Interestingly, we found the shortage of toilets to be a significant contributor to violence among students, as the recess period was too short to permit everyone to have access, and bullying developed.

Evaluators often approach the world with differing perspectives than do specialists in a sector. That is the stance of Garven, Micaro, Rojas, and Wenderoth in their chapter on evaluating gender equality. From the evaluator’s perspective, they consider Universalia’s experience in evaluating gender equality as both a crosscutting theme and as a distinct focus. Six challenges for evaluators are presented, together with ways of responding to them. The challenges range from problems in the conceptualization of gender equality and the nature of gender equality to a lack of frameworks and methodologies for evaluating it. The chapter concludes with suggestions for moving forward for better evaluation of gender equality in the future.

We have one contribution related to the development of evaluation capacity in a geographic region, the Middle East and North Africa (MENA). Doha Abdelhamid explores the state of M&E in the Arab world, which has many differences to other areas. In an interesting and thought-provoking exposition of development in MENA, she reviews the engagement of regional governments and NGOs in global development and then goes on to consider the M&E network architecture. Two international bodies support evaluation: The International Organization for Cooperation in Evaluation, which has organizational members, and the International Development Evaluation Association (IDEAS), which has individual members. Abdelhamid then outlines an ambitious strategy for evaluation capacity development for African and Arab nations. The chapter concludes with required catalytic factors for success and what the international donor community needs to do to support the initiative.

For thirty years, Universalia has tried to link the theory of evaluation with “real life” practice. It is the metaphor of Abelard and Heloise. This edited book is another small attempt to live what we believe. Evaluation generally, and development evaluation more particularly, are applied areas. Theory and practice need to be melded. It is with this belief that we have tried to shape practice and to bring together those thinking about evaluation with those engaged in evaluation. Sometimes they are the same and that is our hope!

SECTION II

*Cross-Sectoral Trends
Influencing International
Evaluation 1980-2010*



CHAPTER TWO

Something Called Capacity: The Evolution of an Idea in International Development

Peter Morgan

The international aid industry behaves in strange ways. It latches on to ideas such as capacity or partnership or ownership or fragility which it claims lie at the heart of development. All future progress, it is said, depends on getting these things right. Then the industry turns around and admits that it cannot define them or sometimes cannot even explain them in ways that most people can understand. Magicians, comedians, and theologians can get away with this kind of stuff, but why does the development business try to do the same thing? What is going on here?

This contribution to the *Universalia Anniversary* volume tries to answer these questions with respect to one of these ideas, that of capacity. It assumes that the evaluation of capacity development will not amount to much if the idea of capacity is itself poorly understood. This article thus looks at the evolution of the idea of capacity as a development concept and why it does (or does not) make sense. It is, to put it charitably, a mixed story. A happy ending is not yet in sight.

We will begin by addressing two contrasting views of the capacity issue and its developmental significance. From the sceptical perspective, the idea of capacity is a ghost concept, striking and somewhat intimidating but without any real form or substance¹. Indeed, many argue that the idea of capacity is much more a symbol or a slogan than a substantive concept. This lingering sense of abstraction can lead to endless, unproductive discussions about definitions and the true meaning of capacity. Even a sense of professional discomfort creeps into the picture. Consultants, for example, hesitate to include “capacity development” on their business cards to avoid the usual awkward questions.

There are good reasons to take this view. The study of capacity does not have the respectability of being an intellectual or academic discipline such as economics or sociology or public administration. The term does not appear much in the vast literature on private sector performance and management. Almost no schools or universities on the planet offer courses or degrees in it. There are no departments, professors, or chairs of capacity studies. No textbook has ever been written to explain it. There are few institutes or research organizations devoted to its development and application as a body of knowledge or practice². Even they cannot bear to use the term on the front door.

¹ “...development management (DM) continues to lack a core identity. The lacuna is both conceptual and operational. Most disciplines are characterized by an agreed-upon set of precepts, principles, and avenues of inquiry that forms the dominant disciplinary paradigm with some challenges to that paradigm at the margin. The DM’s multidisciplinary tent with its fuzzy contours and extensive conceptual and operational reach has resulted in a struggle among competitive concepts with no ‘clear winner’”. (Brinkerhoff and Brinkerhoff, 2010, p. 110).

² An exception would be the International Non-Governmental Organization Training and Research Center (INTRAC) in Oxford, England. Or the Center for Effective Organizations at the University of California.

Capacity, at first glance, appears to have no body of theory or central concepts of its own, for example, the 'market' in economics or the 'institution' in sociology. For the most part, it must borrow its theory and principles from other disciplines such as organizational development or institutional economics or political economy, which have much deeper traditions of working with their own ideas. The concept can also be a puzzle to many outside the English-speaking development community³. Many country participants can be confused about what it all means beyond training and the possibility of more donor funding. There are plenty of capacity sceptics who believe the idea is a bit trivial, and amounts to little more than the latest ploy to keep the aid business going. The central question of these sceptics – what, if anything, does the idea of capacity add to the study and practice of development that is unique and useful? – is a valid one that remains to be addressed.

So, what is the response? What is the other, more positive, view of the concept of capacity? It goes something like this. The sceptics do not quite understand what is happening in development these days. They live in a world of narrow, unidisciplinary solutions and reductionist ways of thinking about the way the world now works. For example, few believe that conventional development theories and policies such as that of economics can claim by themselves to have led to whatever development successes are out there. Something else is going on and that something has to do with the behaviour of complex human systems. In their own way, that is what a new generation of "systems" concepts –

³ The Praxis program at INTRAC at Oxford England has done a series of discussion papers on the idea of capacity from a French and Spanish perspective. For more details, see www.intrac.org.

e.g. ownership⁴, fragility⁵, state building, governance, and yes, capacity – is trying to explain.

The following are key points. First, the issue of definitional diversity is hardly limited to the field of capacity. In practice, every one of these new concepts struggles to define itself in ways that other people will accept. Nor is there anything new about this process. People still disagree about the meaning of effectiveness⁶. The effort to reach agreement on the definition of the idea of innovation has been underway for over 300 years and is still not resolved⁷.

Second, capacity does have a central concept. It is that of ability or competency or skill, especially at the collective level. Two definitions are typical. Capacity is the “ability of people, organizations and society as a whole to manage their affairs successfully (OECD, 2006). Capacity is the “capability of an organization to achieve effectively what it sets out to do” (Alan Fowler, 1995). A slightly different one is that of the Community Development Resource Association (CDRA) in South Africa...”capacity is the “ability to function as a resilient, strategic, and autonomous entity”. As a way of

⁴ “The term local ownership remains vague and undefined in its usage, particularly in policy papers. Even in conceptual frameworks, where the importance of local ownership is highlighted, the concrete meaning or implication of such a guiding principle is barely discussed” Reich, H., (2006), ‘Local ownership in Conflict Transformation Projects: Partnership, Participation or Patronage?’ *Berghof Occasional Paper # 27*

⁵ “Despite this heightened focus on fragility in research and policy, there is no clear consensus on its definition” *Societal Dynamics and Fragility: Engaging Societies in Responding to Fragile Situations*, World Bank Concept Note , Nov. 2009

⁶ “The term ‘effectiveness’ is ubiquitous in international humanitarian aid, diplomacy, development, and defense and security, and although it has many technical meanings as described above, it remains without an agreed-upon definition nor a common understanding of its impact related to complex operations” S.J. Meharg, *Measuring Effectiveness in Complex Operations: What Is Good Enough?* Canadian Defense and Foreign Affairs Institute, October 2009 p. 2

⁷ For a discussion, see Evert Rogers, (1995), *The Diffusion of Innovation*

thinking, it tries to explain how individuals, groups, organizations, countries - indeed any human system – configure or enable themselves to be able do things over time in order to create the kind of future they want. The challenge, of course is to understand how that complex process of capacity development happens in a variety of contexts.

Third, the idea of capacity represents the latest effort in the development community to focus attention on that family of management issues such as implementation, execution, delivery, feasibility, enactment and so on. Most development analyses have traditionally dealt with intent, policy and strategies, i.e. the ‘what should be done’ question. An attention to capacity addresses the ‘how is this to be done?’ and ‘can this be done?’ questions, i.e. how can a policy actually be implemented in real world conditions and deliver the gains intended?

Fourth, donors and countries must deal with development puzzles of increasing complexity such as those related to security or health or climate change. What are needed now to respond to this complexity are systems thinking ideas and techniques. As such, the idea of capacity is one of a group of approaches such as sustainability, reliance, human well-being or governance that attempt to integrate insights from many other ways of thinking such as organizational development, political economy, institutional economics and sociology into a more connected approach to development. Its ‘borrowing’ of concepts is thus a deliberate strategy designed to foster greater coherence.

Finally, the idea of capacity signals the need to get beyond the limitations of the traditional donor-recipient relationship. Donors can signal their willingness to accept a shifting aid relationship in which the country uses or develops its capacity to take the lead. The idea of developing country

capacity can be used as a symbolic or legitimizing device to communicate a new way of approaching development cooperation.

We can begin to see here the potential of the idea of capacity. The severe challenges involved in trying to fulfill that potential. The nascent insights connected to the overall idea are emerging from a different source compared to those of conventional disciplines and ways of thinking. For the most part, it is not coming out of academia or research institutes. It is based on the operational experiences of mid-level donor staff, country participants, consultants, and groups such as the Development Assistance Committee (DAC) of the OECD in Paris. The relevance and vibrancy of the concept of capacity thus depends largely on the effectiveness of their learning processes and their willingness to share experiences. It remains unclear at this point if the energy, commitment, and resources exist in the international development community to give a credible answer to the central question raised by the sceptics and discussed earlier in this chapter.

The Evolution of International Ideas about Capacity

How has the idea of capacity changed over the past decades? This chapter sets out the major shifts in thinking since the 1960s. The aim is to provide readers with a basic guide to the changing perspectives on capacity issues of the international development community. The questions for readers then become the following: given this pattern of evolution, are the resulting insights going to make a significant difference to development practice? Are they going to add up to that

unique contribution? Are they going to remain a somewhat artificial idea created by the international development industry to bolster its own welfare and legitimacy? Finally, what are the implications for its evaluation? Can it actually be measured or at least assessed?

There are two points of explanation. First, these patterns in thinking reflect changes in international thinking and learning about which capacity interventions supposedly work and those that do not. We need to remember that almost all the writing about capacity issues comes from donors and other aid groups. In many instances, country participants may think quite differently about capacity issues. This gap remains a major issue to address over time.

Second, it is important not to overstate the nature and scope of these shifts discussed above. In many cases, older practices of capacity development still have relevance in certain circumstances. It is important for both countries and donors to master the growing range and complexity of capacity development interventions, which are now needed across the spectrum of contexts ranging from middle-income to fragile and post-conflict.

The Shift beyond Knowledge and Technique

A key assumption dating from the 1950s centered on the issue of gaps in knowledge and technique as a constraint to capacity development. Many capacity interventions were designed over the years to help develop more technocratic and functional capabilities in formal organizations, particularly those that donors felt were necessary to help implement and manage their own interventions. The obvious

solution was to transfer expertise from developed to less developed countries mainly through technical assistance usually in the form of the adviser/ counterpart relationship. Capacity in the form of departments, agencies, universities, and civil society organizations could therefore be strengthened or engineered into existence and, at some point, handed over to country participants. Best practice was seen to be a key part of the capacity puzzle.

On the face of it, such an assumption still looks plausible and in a few cases, it can still work. However, it no longer holds up as a foundation of most capacity development strategies for a number of reasons.

First, a good deal of high-income country knowledge and technique did not fit the conditions of low-income countries. Knowledge was treated as an external commodity removed from the bureaucratic and community life of the national participants. To be useful, it needed to be adapted, modified, simplified, and customized mainly by those in the country. Imported knowledge needed to complement and expand upon what the participants already knew. In short, what has come to matter more is the ability of external and country participants to collaborate and to learn and in effect, to co-produce capacity solutions that were not obvious at the outset of the work. Yet, most Technical Assistant (TA) interventions still struggle to organize themselves as learning experiments.

Second, such an approach implicitly focused on weaknesses, gaps, and absences in low-income countries. It assumed that something indigenous and faulty had to be replaced by something imported and functional. Yet, we now know that many country systems have genuine strengths and advantages. In particular, they can have legitimacy and cultural/ social relevance that imports cannot match. The trick then is

to understand these country strengths and build on them in ways that make sense.

Third, the knowledge and technique approach did not address the human intangibles that lie at the heart of any kind of personal and organizational change. We are talking about such things as the ability to change and adapt, to exercise power, to gain legitimacy, to relate, or to inspire. From this perspective, capacity development has turned out to be more of an organic, informal, indigenous process that, if it works, slowly alters the ways in which members of a group or an organization or a society cooperate and work together. Coming to grips with these intangibles has led to the awareness of a range of other issues that go beyond the technical and the functional such as motivation, commitment, human energy, patterns of power and legitimacy, ownership, politics, power, and culture. Thus, in a twist of irony, the so-called hard aspects of capacity development turned out to be the easy part. The so-called soft aspects are now seen as the hard part.

The harder technical capabilities are still essential to enable country organizations to perform and deliver services. People need to be able to manage, for example, complex financial procedures in the public sector. It has become obvious over the last few years that effective capacity depends upon getting complex human systems to work. The challenge for program participants is now to find ways of combining these different aspects – the easy and the hard – in any capacity development intervention.

1. *A shift in the pattern of capacity development interventions from the individual to the wider capacity system*

The focus of many efforts at capacity development for many years was on the micro aspects, i.e. developing the competency and performance of the individual. Over time, the focus expanded to what might be called the meso or the organization. However, over the recent decade or so, many development problems – poverty, climate change, security, and health care – are now seen as complex, system-wide, macro challenges that involve multiple country actors and stakeholders. The resulting efforts to understand the dynamics of these complex systems is changing the way we think about the possibilities for capacity development. Systems thinking about capacity issues, for example, are now becoming more influential. We also now know that the capacity and performance of individuals are shaped in large part by macro factors or the wider institutional and organizational systems of which they are a part. Even micro interventions now need a connectedness to the bigger picture.

Even our image of the composition of a development system is now changing. Earlier efforts at capacity development concentrated on improving the capacity of formal organizations mainly in the public sector, the kind that could be easily recognized as modern – education departments, revenue offices, ministries of planning and so on. But we have come to see, or at least sense, the influence of informal systems and patterns of behaviour on the process of capacity development, e.g. the effects of neo-patrimonial relationships, historical legacies, ethnic networks, coalitions, institutional influences, legitimization, the pattern of hidden incentives at both the political and bureaucratic levels, tacit

flows of information and resources, decisions by unseen actors and many others. Indeed, in some cases, the informal or the 'shadow' side of organizations and systems may be the main repository of capacity and the one most important to change and reform.

2. The rise of complexity, uncertainty and diversity

The spirit behind earlier generations of capacity interventions tended to be confident and optimistic. Generic technical knowledge could be transferred using projects as the main methodology. Linear thinking techniques including detailed planning, results chains, logframes, cause and effect reasoning, specification of inputs and prediction of outputs supported this approach. Not surprisingly, much of this approach fitted well into the practices and behaviours of donor bureaucracies. In recent years, the pressure for demonstrable results and accountability has maintained the enthusiasm for these kinds of mechanical approaches.

Yet it seems clear on the ground that most of these techniques are fast becoming irrelevant and anti-developmental. The shift to macro interventions at the program, sector, and national levels has led to much more complexity given the increase in external and country actors and the influence of political factors. This increased complexity has led, in turn, to the rise of much more uncertainty and lack of participant agreement about the nature and direction of capacity development interventions. Efforts at capacity development now need to depend less on technical analysis especially in advance of implementation. They need much more in the way of facilitation, communication, relationship building, and varied human exchanges. They need less strategic planning. They need more learning, adaptability and flexibility as implementation unfolds.

3. *The influence of the country context*

There is now a growing line of thinking that the country context – or put in systems terms – the country ecosystem – should be the starting point of any serious capacity analysis⁸. We are talking here about the distribution of political and personal power, patterns of interrelationships, incentives, energy, and flows of resources, governing ideas and values, cultural traditions, historical origins. We must also include the influence of these factors on the country participants who must actually do the work. Donors, it is said, need to have much better answers to some basic questions such as ...how do the country's bureaucratic and political systems in which they are intervening actually work in real life? ...Not how they 'should' work, but how 'do' they work⁹. ...What are the strengths? ...What are the traps? ...What are the resulting spaces and opportunities? ...What changes are already underway? ...What are the emerging patterns? ...What can the system likely absorb or accept?

The emerging view is of capacity in the form of organizations and institutions as an emergent complex response to political and social norms and forces in a society. Individuals and groups in a society bargain and negotiate the distribution of power and the allocation of resources. In the process, they both directly and indirectly induce capacity outcomes in the form of accepted or legitimated organizations or systems. From this perspective, participants start with the context and see what possibilities and opportunities exist.

⁸ The OECD Principles for Good International Engagement in Fragile States and Situations

⁹ For one effort, see Goran Hyden, *Why Do Things Happen the Way They do? A Power Analysis of Tanzania*, unpublished memo for SIDA

Alas, again, there is the inconvenient issue of the operational implications. Since the 1980s, the gaining of detailed country knowledge has passed out of fashion in development agencies as both a development and a career strategy. The international development community has increasingly emphasized universal knowledge at the expense of the country. The tendency has been for development knowledge to be quickly aggregated, universalized and homogenized for corporate use ('best practice') rather than customized. Knowledge about technical subjects (e.g. financial management, water resource development) and administrative practices (e.g. performance management, results-based management, and monitoring and evaluation) have become more valued and rewarded.

Because of these trends, most international development agencies are now not structured and/or incentivized and/or trained to do serious contextual analysis compared to say, intelligence agencies. Increasingly, donors face a challenge in coming to grips with the underlying contextual issues in a particular country¹⁰. Such understanding needs to inform and be balanced off against the aggregation and synthesis of knowledge at the corporate level. That, in turn, implies some clear thinking about the overall structure of donor agencies

¹⁰ Nor is this an issue that donors alone find extremely difficult to address. A recent article in the New York Times (January 4th, 2010) quoted the US military intelligence chief in Afghanistan as sharply critical of US intelligence in that country. According to the report, US grasp of the country was "ignorant of local economies and landowners, hazy about who the power brokers are and how they might be influenced ...and disengaged from people in the best position to find the answers" US intelligence was "'unable' to answer fundamental questions about the environment in which US and allied forces operate and the people they seek to persuade". The US intelligence community had a "culture that is strangely oblivious of how little its analytical products, as they now exist, actually influence commanders. We're more than a fingernail deep in our understanding of the environment" A good deal of analysis also exists on the difficulties that private international corporations face in understanding the context in which they work.

and the tension between decentralization and centralization. Contextual knowledge has also been a victim of the 'do more with less' syndrome in donor agencies combined with the belief that programs can shift rapidly to a 'hands-off' self-implementing basis. Shifting that stance will require some effort. Donors have to decide if they wish to develop their own capability for contextual analysis given all their other priorities and the constraints to actually doing it.

4. *The rise of dilemmas and traps*

The engineering view of capacity development that has prevailed for some years implicitly saw it as a linear sequence of cumulative or mutually reinforcing steps – inputs to outputs to outcomes to impact – without much in the way of unintended consequences. We are now beginning to understand it as an inherently contradiction-filled activity riddled with tensions and contradictions. Put another way, almost all efforts at capacity development by both country governments and donors are continually faced with dilemmas and traps¹¹. They are up against competing objectives and trade-offs. Efforts at capacity development have a tendency to be stuck. They fall into organizational swamps from which they cannot escape without outside help. They push for capacity and performance improvement in one direction only to find that things get worse in another. Training, for example, can improve the chances of staff to abandon the organization. The drive for more efficiency can undermine effectiveness and organizational sustainability. In post-conflict states, external help and even control may be

¹¹ The idea of 'traps' is getting fashionable. See Tony Addison, *The Chronic Poverty Report 2008-2009: Escaping Poverty Traps*. Jeff Sachs in his recent Reith Lectures for the BBC focused on four development traps – poor nutrition, debilitating disease, terrible infrastructure, high fertility. Paul Collier also talks about wider development traps in his book, *The Bottom Billion*, chap. 2.

crucial at the beginning of an intervention to stabilize the situation and make citizens believe that something – anything – is being done to make their lives better. Nevertheless, the effort over time creates parallel structures, marginalizes the government, and undermines the very country capacity that is needed to make any kind of sustainable difference, and so on. These capacity dilemmas and traps, as the name implies, also have internal mechanisms that serve to immobilize their victims. The challenge for both governments and donors is to be aware of the dilemmas and traps that await them and to manage them consciously.

5. The dynamics of country and donor ownership and commitment

The past few years have seen a focus on the principle of national or country ownership as one of the keys to capacity development. The assumption has been the following: limiting the intrusiveness and supply-driven practices of donors would help create the space for country actors to claim the driver's seat leading, in turn, to more attention to country priorities, more encouragement of country leadership and motivation and eventually, greater development effectiveness. A supporting assumption has been that of the donor-country relationship as the basic determinant of the strength of country ownership. This connection had become imbalanced given the variation in power, capabilities, and resources between donors and recipients. It needed to be reshaped. Well-intentioned people especially in the countries would then have more space, commitment, and opportunity to do the right thing.

All of this is fine as far as it goes, but we are still at the end of the beginning in understanding and addressing the issues associated with country ownership and its connection to

capacity. Earlier donor assumptions about country motivation and commitment, in retrospect, look somewhat naïve and innocent. People in poor countries, it was thought, needed and wanted to do better, to make progress much like people in rich ones. Commitment and good intentions were inherent in most situations in most countries. What was lacking was the functional and organizational ability or the capacity to do better. Simply put, people were willing but not able. Once more resources were introduced, the internal dynamic of self-improvement and capacity development would take over and induce the changes that were necessary for development.

Not so fast. More intractable issues have now emerged, i.e. the personal, organizational and political dynamics involved in shaping country ownership, most of which got little attention in the earlier discussions centered on the influence of the aid relationship. Dealing with the ownership dynamic has thus turned out to be a good deal more problematic than the aid community had imagined. The international community is still reluctant to admit what is obvious including in our own countries, namely that complex organizational and institutional change in partner countries (in the form of capacity development) is a rat's nest of individuals, groups, agendas and interests contending in various ways for power and resources.

Donors, for example, usually struggle to understand a key aspect of the ownership issue in a post-conflict state, i.e. the attitudes, motivations, and interests of different groups of participants. Capacity development for whom exactly? In what form? And why? And at whose expense? For example, if politics is still conducted in a country on a mainly neopatrimonial basis that by its very nature does not focus on the general welfare and the public interest, what does owner-

ship mean in such a context? And whose ownership? Commitment to what exactly? And why? What happens when the wrong people from a donor perspective – people who benefit from dysfunction and weak capacity – take ownership of an external intervention for their own purposes? Moreover, what happens when country ownership is weak or non-existent on key interventions? At what point do donors bail out as a result? What happens when resistance to an external intervention is a sign of country ownership? Finally, what power, control, and accountability are donors willing to give up or trade-off in order to create the space for countries to exercise more ownership? The challenge in the years ahead will be to design international interventions that connect sensible efforts at capacity development with country sources of motivation and commitment.

A final point might be said about donor ownership (donorship) of capacity development. The usual concern has been about too much donor involvement leading to too much direction and control. An equal concern should be about too little donor ownership, i.e. too little patience for the long haul, too much temptation to try and support capacity development on the cheap and in the short-term and finally, too little inclination to adapt their own policies and procedures to meet these new challenges. What seems to matter is not simply the nature of country ownership but rather the complex interrelationships between these two ownerships – of countries and donors – and the way they interact together to generate capacity outcomes.

6. Capacity development from the demand side

The conventional view from the early decades of capacity building implied something that was engineered from the inside-out through reform strategies, restructuring, organizational expansion, training and so on. From this perspec-

tive, external and country participants worked together to create or improve capacity through deliberate planned interventions focused on needs. In many cases, the design of formal organizations were transferred or imposed – supplied- from the outside through TA or the adoption of international best practice. Contextual factors were seen mainly as constraints or conditions whose influence needs to be taken into account.

We now know more about the impact of capacity being demanded as well as supplied. The assumption here is that capacity in the form of a formal organization or even a government delivery system will not by themselves retain its responsiveness and relevance. They need to be pressured and made accountable by those they are in business to serve. At the root of the problem is the inability of the consumers of these public services in low-income countries to invoke the exit option in the same way that customers of private firms can refuse to buy goods and services.

This process of capacity development through demanding can happen in a number of ways. The long route is through democratization that, in theory, gives voters the ability to pressure the government to perform. The short route is equipping citizen consumers to use things such as citizen report cards as a means of holding public organizations more accountable. The market route is to put citizens in the position of private sector customers who can exit from the provision of services in health or education or water¹². Whatever the route, the assumption is that a combination of

¹² For capacity building programs that use the 'market' route, see Jacqueline Novogratz, (2009), *The Blue Sweater: Bridging The Gap Between Rich and Poor in an Interconnected World* . Also Muhammed Yunus, *Building Social Business: The New Kind of Capitalism that Serves Humanity's Most Pressing Needs*, 2010.

both demand and supply will energize the process of capacity development.

These different views of capacity development point to the growing need for participants to think through the theory of change that they are using. Earlier approaches to capacity development relied on the principles of planned and mechanical approaches to change. Such interventions tended to be structured around the budgeting and planning cycles of donors, e.g. three to five years. Perspectives that are more recent pay more attention to political economy, learning, the creation, dissemination, and use of knowledge, historical influences, the effects of institutional frameworks and the evolution of complex systems. In addition, the perspective on the time issue has changed. Quick wins are now fashionable as a way of generating and maintaining support in the short term. At the other end of the spectrum, some capacity development processes are now seen as part of complex historical patterns of evolutionary change that can take decades to unfold.

7. The growing strategic importance of capacity issues

In the 1960s and 70s, the prospect of massive poverty in the South was seen as a major threat to the long-term strategic interests of countries in the North. Ultimately, global peace could not be sustained, it was thought, in the absence of the economic development of low-income countries. This argument may still be true but the discussion has shifted. Now, the focus is more immediate and short term. In an increasingly interconnected world in which individuals and small groups can have a global reach, the capabilities of countries, groups, and individuals in the South now matters critically to those in the North. The capabilities of the Yemini security services were unknown in 2008. By the end of 2009, it had become a subject of regular discussion in American

newspapers. In the early 1990s, the effectiveness of the Pakistani education sector was the subject mostly of World Bank social sector supervision reports. Now, its dysfunction and its relationship to radical madras is a preoccupation of western governments. The state of the Indonesian Ministry of Health and its capacity to contain outbreaks of bird flu that could spread around the world catches the attention of anxious citizens in Australia and The Netherlands. National capacity is no longer just about program implementation. Its absence can now have global repercussions and be the subject of debate in the UN Security Council. Other global public goods such as the capacity to address climate change are next on the agenda.

Put another way, capacity development is now about helping people in partner countries put in place the institutions and organizations, both formal and informal that enable them – or not – to make progress. This is the general theme put forward by Amartya Sen in his writings on development as capability expansion. In this sense, capacity should now be seen as a development goal in its own right as a development end and not just an operational means. It is now about the “what” of development and not just the “how”.

8. The evolution of the external support role

Creating the role of donors and other external interveners in supporting capacity development has become much more complex from two perspectives. As mentioned earlier, donors began by adopting a more intrusive role in terms of building capacity. TA could help to erect or strengthen formal structures, which could then be taken over by country staff. Inherent in this approach was the tension between direct task accomplishment – getting the job done – and indirect support to capacity development.

However, four emerging trends have become evident. First, the trend is for external interventions to be less and less intrusive, to leave more space for country participants to control their own activities. Debates about country ownership, budgetary support and the evolving positioning of technical assistance are all part of this evolving pattern. As is the growing number of talented country participants able to address capacity issues on their own.

Second, the accelerating complexity of development has created many more possible support points (as opposed to entry points). External actors can now think about providing facilitation, political support, leadership training, market access, and many other varieties of support designed to have some onward capacity development effect. Third, most recent international agreements such as the Paris Declaration and the Millennium Development Goals now encourage donors to harmonize and coordinate under country leadership and through country structures. The implications for the development of capacity under these more harmonized and collective approaches remains to be seen. Finally, we are seeing the rise of alternate sources of funding for capacity development including new development agencies from countries such as India or Brazil, mega philanthropies such as the Gates Foundation and the private sector entering into social sectors. Most of these new actors will bring new approaches to the capacity issue.

9. The odd combination of enthusiasm and ambivalence

A cautionary point needs to be introduced. The international development community – donors, countries, international NGOs, -- has tended to show a good deal of tacit ambivalence about the capacity issue. On the one hand, capacity has

variously been labelled as the missing link, a core function, a pillar, a strategic priority and other similar terms. Agreements such as the Accra Agreement in 2009 make much of its importance. It is hard to credit this commitment by looking at both activities on the ground and at donor investments in their own capacity to support capacity development. The cold reality is that the pressures and incentives that shape donor behaviour do not combine to make capacity development a real priority along the lines of say, RBM, or poverty.

Why is this? Because in the current aid environment of scepticism, impatience and the need to demonstrate short-term results, the capacity issue simply comes with too many liabilities. It is hard to figure out and understand what specifically to do. Devising strategies to implement complex change, for example, remains a puzzle. The issue is littered with intangibles that cannot be measured or claimed. Its elements cannot easily be stuffed into logframe categories. Risks are high. Success rates appear to remain low. The main benefits of capacity development are suspected to be long-term or past the point when the credit can be claimed by any current participants. Moreover, the enthusiasm of country participants themselves for programs of dramatic change can be limited.

Part of the ambivalence is long-standing. Capacity development and its onward connection to performance and results are basically about execution or implementation. For different reasons, neither donors nor country governments have given these issues much sustained attention over the years. Both groups are rewarded for and hence more interested in generating policies, prescriptions, strategies, intent and the commitment of funds. Capacity is not a priority, for example, for the Millennium Development

Goals. It has never been the subject of a World Development Report or a UNDP Human Development Report. It is usually not addressed in any depth in exercises such as PRSPs. And so on.

The bureaucratic dynamics of donors are another part of this pattern. The capacity issue arouses little or no interest in the public of donor countries, which understandably responds much more to issues to do with education or health or poverty. Not surprisingly, political leaders and senior bureaucrats in donor countries are usually uninterested in the issue except in connection with security issues¹³. Capacity development also has no domestic groups in donor countries lobbying for its inclusion in aid programs, as is the case with gender or human rights or the environment. In practice, its main supporters are usually middle managers in donor agencies who are accountable for the implementation and delivery aspects of aid programs. In political terms, the influence of the capacity issue and lobby on donor funding and domestic legitimacy is marginal. Bottom line: the ownership of the capacity issue by the international development community is uneven at best.

Some of the same ambivalence is evident at the country level. Symbolically, capacity development is felt to be a good thing if it contributes to country control and sovereignty. However, some of the same baggage remains. Country participants are tired of the same old technical assistance and the endless stream of imported practices and models in the form of capacity 'assessments', evaluations and reporting structures. Countries increasingly have other options such as China or the private sector, other sources of knowledge such

¹³ The Canadian Minister for International Development has reportedly said that she does not want to see the term 'capacity development' included in any memos sent to her for approval.

as the Internet and more sources of domestic or diaspora expertise. In addition, many country officials tire of the demands for the organizational transformation of their structures from donor countries whose own inability to transform their own systems is evident.

The Way Forward

We are confronted once again by a paradox. Capacity is needed by all societies to make progress. Individuals, groups, and organizations need to be able to contribute, to make a difference, to perform in a way that benefits the people they serve. This basic point remains true for all countries. Yet, the symbolic enthusiasm of the past years surrounding the capacity issue has waned. So what is it that capacity advocates can highlight in an effort to keep this aspect of development front and center?

The following three things are suggested:

- First, capacity advocates need to bring a positive vision to the development debate about the value of capacity as a key element of human progress¹⁴. In the words of Brinkerhoff and Brinkerhoff, “development managers should come to the journey with a good dose of courage, vision and a pragmatic sense of hope”.¹⁵

¹⁴ Abouassi, K. “International development management through a southern lens” *Public Administration and Development*, Vol. 30, 2010

¹⁵ Brinkerhoff, J.M., and Brinkerhoff, D.W., *International Development Management: A Northern Perspective*” *Public Administration and Development*, vol. 30, 2010, P. 113

- Second, the core concept of capacity was described earlier as that of collective ability. Capacity advocates should be able bring to any discussion insights about how collective abilities form in different systems and at different levels, and what those abilities would look like and how they would contribute to implementation and performance. To do this, our collective ability to research, to learn and to illuminate needs to be maintained. This volume by Universalialia is an example of this process at work.
- Third, the theme of this Universalialia collection is that of shaping international evaluation. The development community still lacks well-conceived tested methods of evaluating capacity development that are useful and accessible to practitioners in both development agencies and partner countries. Better ways of assessing the right things in the right way for the right reasons are urgently needed. This volume will make a contribution to that issue.

CHAPTER THREE

Results-Based Management

Ken Stephenson

This chapter examines the thinking behind results-based management (RBM) and the reasons for its widespread failure in practice. In this chapter, we argue that one of the main reasons for this failure is confusion around the intended uses and intended users of RBM. While RBM is typically thought of as existing at one level, in reality it usually describes a relationship between two or more levels, which affects how performance information is used and by whom it is used. The accountability relationship between levels and the learning requirements of each level in this relationship inform the nature of the RBM approach that should take place in that context. A lack of clarity on these specifics makes it easy for RBM to serve a purpose opposite to that which it is intended, to prevent management for results rather than encourage it. In this chapter, we present three types of RBM that can be used in the real world, depending on the use intended.

Background

In the sphere of international development, RBM was designed in response to the realization that much development programming reported primarily on outputs produced

rather than outcomes achieved. Reports on program performance might proclaim, for example, that a certain number of wells had been built as part of a rural development project, without reporting on how those wells might have influenced, for example, the incidence of water-borne disease in the targeted communities. Since bilateral and multilateral agencies would typically hand out funds to executing agencies to carry out development programming locally, donors themselves would have no knowledge of the end results of development programs beyond what was contained in the reports of program implementers and monitors. Donors were therefore required to make decisions on programming in the absence of any knowledge of how these programs affected the lives of those they were aiming to help.

This lack of attention towards outcomes also affected public reaction towards development. Citizens of developed countries, whose taxes paid for development programming, heard stories about projects gone wrong, such as tractors that lay abandoned and rusting in farmers' fields. These citizens then began demanding greater accountability for the end results of development interventions. Bilateral and multilateral agencies were under great pressure either to show they were making a positive difference, or to stop wasting taxpayers' money.

The logical framework, or logframe, was introduced by USAID consultants in 1969 within this context. Originating from an engineering background, and influenced by similar movements in the private sector, the logical framework analysis attempted to establish a basis for holding programs accountable for achieving what they set out to do. Although somewhat complicated to look at, the idea behind the logframe was simple: it asked of a program what it intended

to achieve, and how it would know when it got there. Other planning innovations came later, including the performance measurement framework (PMF), all with similar aims.

It is no secret, however, that RBM has not worked in practice. The 2008 Review of Results-Based Management at the United Nations, by the UN's Office of Internal Oversight Services, concludes "Results-based management at the United Nations has been an administrative chore of little value to accountability and decision-making". Similar reviews in other organizations have scarcely been more positive. RBM tools and principles often overlaid an existing set of norms, rules, and practices within the organization more geared towards compliance than to learning and risk-taking. This often ended up meaning the term "accountability for results" was simply a new term for the traditional requirement for compliance.

What is RBM?

Later in this chapter, we present an outline of what RBM has become in practice. First, one must be clear on the ideal of RBM in theory.

RBM has two sides: information generation and use. Information generation involves collecting data on the performance of the program, initiative, or organization and analyzing that data to arrive at a greater understanding of what results are being achieved. The second side is simply the use of that information, which relates to the purpose of doing RBM. The two purposes of RBM – the two uses of performance information – are learning and accountability. Learning is the use of information to inform decision-making, set policy, or in some other way improve the

program, initiative, or organization. Accountability means demonstrating the achievement of results to somebody to which one reports¹⁶.

Results-based management must be distinguished from other practices to which it bears some similarities. Within development, for example, monitoring is a familiar component of practically any project or program. Donor agencies monitor implementation of programs on the ground to ensure implementation follows agreed plans. Many people view RBM as identical to “monitoring for results” – the traditional task of program monitoring, with a particular focus on results. This definition can be too narrow because it tends to over-emphasize the accountability component and under-emphasize the learning potential at program level. It over-emphasizes the information-generation side of RBM at the sake of use of that information.

Others may view RBM in development as a program management tool. Some agencies have manuals on RBM that do not mention accountability once. Rather, this side argues, RBM is a way for program managers to manage their own programs better, if only they can be forced to adopt the practice. This definition is not ideal either because RBM as it is practiced today is often little used by program managers, who tend to view it primarily as an accountability mechanism.

In our definition, RBM encompasses “monitoring for results”, evaluation, and any other activity designed to improve accountability and learning in a programming context based on a focus on results. It is not a tool nor a

¹⁶ This is a narrow definition of accountability, which has been the primary focus for much of the literature on RBM. A broader definition of accountability would include other stakeholders beyond those to whom one has an official reporting relationship, including intended beneficiaries, partners, and others.

methodology, nor a discipline unto itself, but rather a new perspective on what we already do. RBM is only given a name at all because of a lack of adequate results focus in traditional management. In 20 years, if properly implemented, there will no longer be the term “RBM”; rather, it will be called “the right way to do things”.

The Local Level and the Central Level

The literature on RBM, traditionally, describes it as existing at one level only: the level of the program (or project). In recent years, there has been growing realization that RBM also has its place at a strategic level within organizations, such as bilateral and multilateral aid agencies, in attempting to demonstrate results to their donors and to set strategy and policy at an organization-wide level.

In reality, however, it is often more accurate to think of RBM as a process informing a relationship between two or more levels. We will use the terms “local level” and “central level” to describe two levels whose relationship is defined in part by RBM. These can refer to practically any two levels seen in real-life programming. In a donor relationship, the central level is the donor, and the local level the recipient of funds. In describing external accountability relationships of an organization, the local level may be the organization itself and the central level its board, donor committee, or similar body. Within an organization, the local level may be the field office in country, and the central level is the organization’s headquarters office. Similarly, across partner organizations, one may think of the local level as program management units and the central level as the executive office within the lead organization.

In all cases, the local level is closer to the implementation of program activities, while the central level oversees in a strategic, policy-setting, and/or coordinating role. The central level may also have its own relationships with others – for example, there are likely to be bodies to which it is accountable for demonstration of results.

The purpose of distinguishing between levels is to understand that the particular accountability relationship between the local and central levels, along with the learning requirements at both levels, shape RBM into what it should be in any particular context.

In international development, one relationship that is very common is one of donor and recipient. In the language of development, there are no longer donors and recipients per se; we are all now “partners”, a term that deliberately gives the impression of equality. This equality is especially important in the domain of development, where programs connect developing nations to developed ones, and emphasizing an accountability relationship between the two may raise an uncomfortable suspicion of neo-colonialism.

Within this context, it is sometimes difficult to examine issues of accountability. Well-established systems, such as audit, go unquestioned, but relative newcomers such as RBM face criticism from many as an example of the tightening grip of control over the affairs of developing countries, at a time when the world is gradually coming to the consensus that developing countries should have ownership over their own development.

This presents difficulties in the development of a theory of RBM. Agencies are often reluctant to utter the word “accountability”, while this is very much what RBM comes to be about in practice. The inability to express honestly what

RBM is and is not has, in my opinion, allowed the entire concept to become muddled, and corrupted in practice, into what I call “False RBM”.

False RBM

False RBM is the type of RBM most commonly found in practice, and it is RBM in name only. It is structured in a way that it is useless for accountability purposes, and actively discourages any real management for results. The overall aim of False RBM is to project the illusion of a results focus while still maintaining tight, centralized control based on compliance. Aspects of RBM that do not threaten this control (such as development of planning tools) are encouraged, while aspects that do threaten this control (such as use of these planning tools in implementing the program) are actively prevented.

False RBM may be identified by some or all of the following characteristics:

- RBM guidelines that lack clarity or contain contradictory statements about the end users, or the end use, of RBM information.
- RBM guidelines that place excessive emphasis on development of planning tools (logic models, performance measurement frameworks, etc.) and little to no emphasis on the use of these tools.
- An authority structure that places decision-making authority heavily in the hands of the central level, combined with an RBM relationship that requires the local level to demonstrate and manage for results.

- Performance reports whose narrative is heavily activity- or output-focused. If outcome information is included, it is in the form of “objectively verifiable indicators” which are not analyzed and bear little or no connection to the report narrative.
- Performance reports that contain uniformly good news (i.e. no opportunities to learn from previous experience).
- No adjustments made to program design or implementation after each RBM cycle. Formal or informal restrictions in place that prohibit these adjustments.

RBM becomes False RBM because of certain pressures that hinder both the generation of reliable performance information and the use of that information. Unreliable performance information comes about when the incentives for good news stories trump any incentives for transparency or accuracy. In a worldwide development system, where implementing agencies need to ensure continued funding and their donors want to show these agencies in the best light to their donors and the public, no one is playing the watchdog: it is not in the interest of those involved to publicize any bad news. The tendency towards non-use of performance information comes not only from the recognition that the information is unreliable for the reasons mentioned above, but also from traditional authoritarian organizational structures driven by risk aversion. Many organizations feel it preferable to centralize decision-making rather than risk any actions by local-level managers that might draw criticism from its stakeholders or the public, thus making the use of performance information at the program level impossible. Because of these pressures, RBM systems tend to result in unreliable information that goes unused.

Accountability and Learning in False RBM

Accountability under False RBM can take one of two forms: accountability for RBM processes or accountability to fabricate results.

Accountability for RBM processes simply requires that certain bureaucratic paperwork be filled out and nothing more. In addition to traditional compliance, this type of accountability adds the requirement to develop a logic model, performance measurement framework, or other “RBM” planning tool. Nothing is learned, no real changes are made to programming or governance other than the requirement for additional paperwork.

Accountability to fabricate results requires adherence to all aspects of a logframe or other planning document, which often spells out in explicit detail the activities and outputs of the program. Little deviation is allowed from this initial blueprint. Program implementers are “accountable” to demonstrate results, but not given the authority to make any changes necessary to achieve real results. One of two consequences occur: either the original design was perfect and results are achieved with no changes needed to the original design – in which case, there was no need for results management to begin with – or else performance information reveals that some changes are required to programming, which are discouraged within False RBM. In this case, the implementer is hindered from achieving results, but accountable for demonstrating results – in other words, is accountable for fabricating them.

Most programs begin with a logframe these days, and in most cases, the original logframe is unrealistic. The expected

results and targets set for programs tend to be wildly over-optimistic if the logframe is developed during the proposal stage of a competitive bid process, as is often the norm. Agencies who compete for funding often find they need to inflate expected results in order to gain an edge over others. However, this over-optimism is common even in the absence of this competitive bid environment, for instance due to the natural enthusiasm of program stakeholders, or because of a lack of understanding in the conceptual stage about contextual factors that act against the program.

Even if the original logframe was accurate, it soon requires alteration to adapt to the changing environment of programming and the lessons learned from each preceding cycle of performance measurement. Most development programming finds itself in a complex environment. Its theory of change is nonlinear: the program interacts with its environment in a dynamic fashion as it is being implemented. This, in fact, is the very essence of the “learning” component of RBM. In this context, a logframe written in the past soon loses relevance.

Program managers may therefore request a change be made to the logframe. Funders are often reluctant to allow these changes, however, since they amount to a renegeing of contractually agreed terms in the world of accountability for results. It is often difficult for funders to distinguish legitimate requests for changes (due to an unrealistic logframe or changes in the programming environment) from more suspect requests (because actual results are not up to expectations). However, although managers are often tied to a logframe that does not reflect the program, and must demonstrate unrealistic or irrelevant “results”, they are still able to satisfy these irrational requirements by fabricating results.

Critics of RBM and the logframe approach have noted in the past that performance can always be demonstrated. This for two reasons: firstly, outcomes are hard to attribute to program activities and outputs; and secondly, outcomes are hard to measure. The attribution problem tells us that the further away one goes on the logic model from program outputs, the more external factors influence results achievement, and therefore the more difficult it is for a program to claim credit for achievement of those results. In the domain of international development, the influence of the program relative to other outside influences is so small that programs typically cannot reliably influence even immediate outcomes. In this case, program managers often count any events that sound like positive results as achievements of their program, regardless of any real evidence linking the event to the program¹⁷.

Another reason results can be fabricated is due to the difficulty in measuring achievement of outcomes. It is difficult to measure many “soft” outcomes that are more and more common in development, such as increased organizational capacity, social inclusion, or improved political will. Exacerbating these measurement problems are the difficult environments in which many programs operate, where structured and systematic data collection is prohibitively resource-intensive, and few third-party data are available.

False RBM therefore creates an internal contradiction by simultaneously reinforcing and denying traditional compli-

¹⁷ John Mayne suggests a contribution analysis is preferable to an attribution analysis in solution to this problem. However, Mayne’s description of the contribution analysis places more emphasis on accountability than on learning as a purpose for RBM, and is therefore closest to the RBM type called Competitive RBM in this chapter. Mayne’s contribution analysis is too onerous to be timely and useful for local level learning and management, for which we refer to in this chapter as Improvement-Focused RBM.

ance-based management. In theory, the traditional accountability requirements of compliance to outputs have been replaced by accountability for outcomes. In practice, paradoxically, the requirement to develop logframes at planning stage has reinforced central authority over decision-making, by binding program implementers to carrying out precisely those activities and outputs described in the logframe and other planning documents. At the same time that central-level decision-making authority is being reinforced, the local level is forced to account for “its” results.

As well as restricting program management, False RBM typically is inactive at the central level. Although in many cases management for results at the local level is impossible because of a centralization of decision-making authority, performance information could still be used to influence decision-making at the central level. However, in our experience with international donor organizations, very few are able to confidently state what results they are achieving at the organization-wide level. Partially, this is due to the newness of applying RBM to the central level, the lack of awareness about how this might be done and of how it differs from program RBM.

Measurement and Analysis within False RBM

In most guidance documents on RBM, data collection revolves around the performance indicator. Data collected are typically numerical or categorical – there traditionally cannot be longer narratives of description about what is happening on the ground.

In evaluation, as in other research disciplines, a “paradigm war” raged for several decades between proponents of quantitative methods and proponents of qualitative methods. A majority of evaluation thinkers now believe that mixed methods, which combine both quantitative and qualitative methods, can offer greater benefits to a study than either alone.

Such a war, or at least discussion, is now overdue in the field of RBM. RBM has historically tended towards quantitative measures, not out of epistemological concerns, as was the case in evaluation, but for purposes of accountability. Accountability requires clear and verifiable yes/no answers to the question “were results achieved?” Of course, these “objectively verifiable indicators” (OVIs) are not necessarily quantitative: they can involve categories such as “partially met” and “fully met”. In other words, they must fall into one of the four levels of scientific measurement (nominal, ordinal, interval, or ratio). Traditionally, they cannot be long paragraphs of texts.

Our recent review of the RBM guidelines of five bilateral development agencies in North America and Europe reveals what may be the current wisdom on performance indicators. Although all five texts examined allowed for both quantitative and qualitative indicators, the definitions in the texts of qualitative indicators corresponded to one of the four levels of measurement above: they were defined either as categories of things or as the quantitative measurement of perceptions or quality. Of the five texts, only two recognized that other information beyond performance indicators is necessary to understand the results of a program, and these two only made brief references to this fact. According to these texts, the primary, if not the only, source of information a

program should collect on results is in the form of numbers or categories of things.

In practice, and in defiance of these guidelines, more and more programs and agencies are making use of “thick description” indicators: qualitative indicators whose data are comprised of long text, broadening the examination of program consequences beyond the limited focus on expected results, and situating these within the contexts in which they occur. They implicitly understand that what they need to manage well is not a count of things, but rather a deeper understanding of the situation on the ground.

For example, over the past few years Universalialia worked with a multilateral development organization on aggregating information on the organization’s results to the global level. We spent a lot of time working together with the heads of this organization in thinking through what information would be useful to learn from what they are doing on the ground in order to make better-informed decisions at the strategic level. In the end, there were very few quantitative indicators: mostly field offices were asked for detailed descriptions on what came about in consequence of their programming, and the primary challenge then came in figuring out how to aggregate this information.

Although useful for management, qualitative text cannot be used for summative accountability purposes. Detailed description does not provide an unambiguous answer to whether or not contractual obligations have been fulfilled. For this reason, the purpose of RBM must be clear from the outset.

Overview of Three Types of RBM

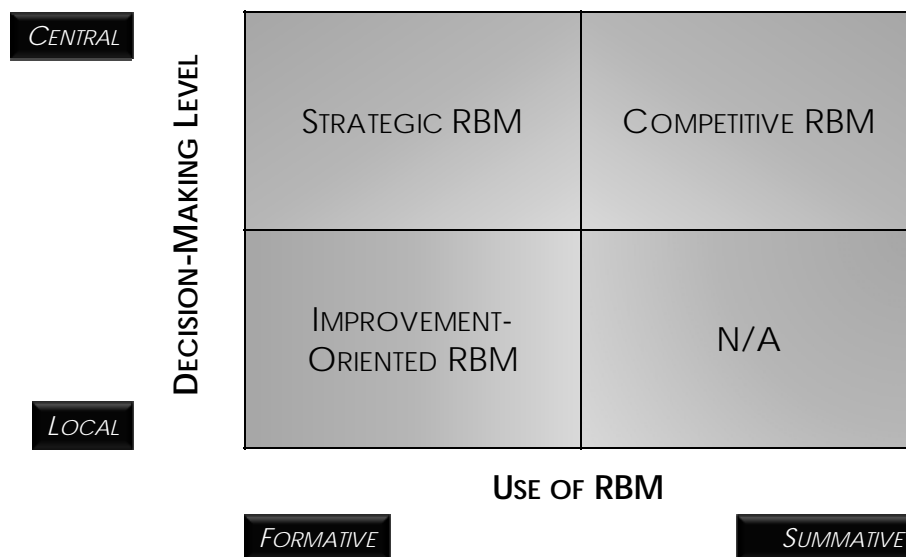
False RBM does not work because it is designed not to be used. So what to replace it with? We need a system which:

- Has use in mind from the beginning. The system needs to be designed to maximize the likelihood it will be used. Ideally, incentives for use should be designed into the system itself.
- Tailors the information generated to the appropriate decision-making level.

The following chart shows the two factors above in a grid format. Decision-making level indicates what level the RBM system is designed to help: the central level or the local level (e.g. the donor or the recipient). The use refers to whether results-focused reporting is forward-looking or backwards looking. Formative means performance information is generated for the purposes of learning, while summative means performance information is generated for the purposes of accountability. It is important to note that accountability plays a role in all three RBM types; however, in the “formative” quadrants, the nature of the accountability relationship does not involve demonstrating results achievement.

The chart shows three types of RBM: Strategic, Improvement-Oriented, and Competitive RBM. Each of these types is carried out for different reasons, to map out different relationships between at least two levels. Each of these has its own requirements.

USE OF RESULTS-BASED MANAGEMENT



The purpose of distinguishing these three types is to be clear on what it is one wishes to accomplish with a particular RBM relationship. One may incorporate all three RBM types above to some extent. However, distinguishing the types helps to clarify that each type needs its own indicators and different processes for validating, analyzing, and using the information generated. Understanding these three different types helps to achieve clarity and consensus among all actors about what you are doing RBM for, and therefore how you go about doing it.

The objectively verifiable indicators generated for Competitive RBM, used to judge the success of a program, are more or less useless for program management, for example, for which more timely and more context-based information is required (for which we use Improvement-Focused RBM). Information generated at the local level for program management, similarly, cannot be aggregated to the strategic level without prior planning, due to inconsistencies across

local level units. This involves separate activities from simple program-level RBM and is thus granted its own type: Strategic RBM.

THREE TYPES OF RBM COMPARED AND CONTRASTED

CRITERIA	STRATEGIC RBM	COMPETITIVE RBM	IMPROVEMENT-FOCUSED RBM
ACCOUNTABILITIES OF LOCAL LEVEL TO CENTRAL LEVEL:	FOR INFORMATION	FOR RESULTS	FOR LEARNING
PERFORMANCE INFORMATION IS USED FOR:	STRATEGIC MANAGEMENT, COORDINATION OR POLICY SETTING	SUMMATIVE DECISION-MAKING	PROGRAM MANAGEMENT
TYPE OF DATA GENERATED:	DATA MEASURED CONSISTENTLY ACROSS PROGRAMS ARE AGGREGATED TO STRATEGIC LEVEL	OBJECTIVELY VERIFIABLE INDICATORS, SUBJECT TO INDEPENDENT EVALUATION AND/OR AUDIT	ANY SYSTEMATICALLY GENERATED INFORMATION THAT IS DEEMED USEFUL TO PROGRAM MANAGEMENT
RESULTS LEVELS OF GREATEST INTEREST:	MEDIUM-TERM TO LONG-TERM OUTCOMES	SHORT-TERM TO MEDIUM-TERM OUTCOMES	SHORT-TERM OUTCOMES
PROS:	CAN MAINTAIN CENTRALIZED CONTROL OVER PROGRAMS; CAN ADD STRATEGIC VALUE TO PROGRAM IMPLEMENTATION	CAN ENSURE GRANTEE PERFORMANCE; CAN MAKE MORE INFORMED GRANTING DECISIONS	CAN ENCOURAGE PROGRAMS TO CONTINUOUSLY IMPROVE
CONS	RISK OF LOSING RELEVANCE AT THE LOCAL LEVEL	LIMITED ABILITIES TO LEARN ACROSS ORGANIZATIONS PERFORMANCE INFORMATION IS NOT USEFUL FOR PROGRAM MANAGEMENT	MORE DIFFICULT TO COMPARE ACROSS PROGRAMS FOR EFFECTIVENESS

None of the above RBM types requires demonstration of results each year. Competitive RBM is the only one of the three in which demonstration of results is even an issue; and for that, we stress that several years are necessary for most programs to understand the impact they are having.

In the following sections, each of these three types of RBM are described.

Strategic RBM

Strategic RBM is RBM done for the purposes of decision-making or coordination at the central level. The local level is accountable to the central level not for demonstrating results, but simply for providing the information needed, which is identified in advance and is aggregated to provide a systematically-generated picture of progress against strategy. Unlike Competitive RBM, the information generated from Strategic RBM is not used to judge local level units in any way. Rather, the purpose is formative: it attempts to identify performance against strategy in order to make changes to it on an ongoing basis. In other words, the information generated from this results aggregation is used to inform the new strategic plan, set policy, or otherwise “steer the ship”.

Since RBM was traditionally thought of at the level of the program, many organizations have had difficulty understanding what shape RBM would take at the organization-wide level. Often they attempted to duplicate what was done at the local level, creating a logframe for the organization, measuring results in a similar way a program would, and using (or not using) performance information in the way they were accustomed to do for a program. However, they have had little success, and in our experience, almost no organizations above the level of a community-based organization are able to systematically say what results they are achieving.

Managing against strategy requires information on performance, just like managing at the program level, but unlike programs, strategic managers usually do not measure performance directly, they aggregate data from various offices, implementation sites or partners implementing the strategy at the local level. Because of this, a strategic manager's ability to get information is reliant on the information being collected by program implementers or others. It is in the interests of this strategic manager to ensure the right information is being collected, that it is being collected in the same way in all areas, that no areas are being overlooked, and that no duplicates are being reported (such as two offices reporting the same result achievement). Regardless of whether an organization's strategy is developed from the top down or bottom up, the process of planning for information aggregation must be mandated from the central level.

Strategic RBM is focused on a different kind of result from what would be measured for program-level RBM. Information of greatest interest would be what might be considered medium- to long-term outcome levels for a program: those results that do not appear as immediate effects of program outputs, but rather show up later, as only an indirect consequence of these outputs. For example, a development organization that targets HIV/AIDS prevention as one of its key strategic priorities may carry out a number of programs, including a public education campaign, a school health program, and clinics offering testing, counselling and care. For each individual program, it may not make sense to measure the incidence of HIV/AIDS in the population, since it is difficult to attribute this to any individual program. However, at the strategic level, it may be feasible that HIV/AIDS incidence is measurably affected by the combination of programs implemented, and that the change in this incidence might tell the organization something about the

success of this strategic priority. Therefore, the central level should request the local level to collect data on this, even though it cannot be used either for formative purposes by the local level, or for judging the success of the local level in implementing its programming.

What should be measured in Strategic RBM is not a full logic model but the outcomes of interest and a small set of factors that may influence outcome achievement. Both the outcomes and the factors need to be elaborated when defining the strategy. A factor may be an output, process, level of organizational “presence”, or an external influence. The local level should not necessarily be forced to adhere to these mechanisms. For example, if the factor is a certain output, different implementers at the local level may be free to decide whether or not to produce that output. Then data on the output in question and outcomes are collected and aggregated to the central level, and differences are observed between those who implemented the output and those who did not.

Unlike RBM at the program level, Strategic RBM may also examine the internal performance of an organization. An organizational strategy typically includes expected outcomes related to internal, structural developments not related to program results. For example, an organization may set as a strategic priority the development of internal capacities on evaluation. Many organizations, for this reason, have two results frameworks at the organization-wide level: a “management results framework” that focuses internally, and a “development results framework” that examines programmatic performance. Since information on the former will not be collected by program-level performance measurement, the strategic manager will have to find sources for these data in other parts of the organization.

Typically, results aggregation happens at several levels. Some data may be collected at the project level, others at the country level, and still others at the regional or global level for multinational organizations. Data often need to be disaggregated to different levels as well. One key challenge for multinational development organizations is to enable disaggregation to the country level, to ensure the extent of alignment to national level goals and priorities, in adherence to Paris Declaration principles.

Strategic RBM can be done in a centralized or decentralized environment. In a centralized environment, where decision-making authorities rest at the central level, Strategic RBM is of particular importance as it is the only one of the three types of RBM that can be practiced. There can only be management for results at the local level to the extent that there is authority to manage. In a decentralized organization, however, Strategic RBM can accomplish more than what is described above. It can also be used to coordinate lessons learned across local level units. The central level would therefore act as a secretariat for learning for the local level. This is most effective when combined with the Implementation-Focused RBM, below.

Strategic RBM does not in itself create incentives to use performance information. To do so, an organization and its board or other accountable body may set up a system of Improvement-Focused RBM as described below.

Competitive RBM

Competitive RBM is used to demonstrate performance. It is useful in granting relationships, where a donor (such as a bilateral or multilateral agency) provides funding to a

recipient (such as a national or international NGO) to implement international development programming. Competitive RBM outlines the outcomes to be achieved in the contract between the central level and the local level, and holds the local level accountable for achieving these outcomes. The extent to which outcomes have been achieved at the end of the period influences decisions made by the central level on resource allocation in future periods, including whether to continue the funding arrangement, find a new recipient, or terminate the program.

The result levels of greatest interest for Competitive RBM may be the short- to medium-term outcome levels: those results that either occur immediately or not long after program outputs. Long-term outcomes are often too far removed from the program's sphere of influence to act as a basis for summary judgment of a program's worth. For example, for a training program, a donor may hold the implementing agency accountable for ensuring that the trainees' level of knowledge is increased, and that they demonstrate behaviour commensurate with this training. The donor should allow some flexibility on the details of the training (activities and outputs), and should not hold the implementing agency accountable for longer-term results such as organizational changes, which are subject to many other factors.

Ideally, Competitive RBM places requirements on the achievement of outcomes but leaves some room for adjustment in terms of activities to be undertaken and outputs produced. If you require accountability for both outputs and outcomes, it constrains the ability of programs to adjust to what it has learned from performance information. RBM requires some element of risk-taking to reach desired outcomes; if your organization is not comfortable with

allowing those types of risk, it should not pretend it is doing RBM, at least not at that level.

In order for the central level to clearly communicate, its expectations at the outset to the recipient, and judge the success of the recipient in meeting those expectations at the end of the period, objectively verifiable indicators are used. However, unlike False RBM, much attention will be paid to data validation and analysis. The governing bodies should have active involvement in the development of the indicators and targets to ensure relevance and appropriateness, and should ideally audit the data collection process to strengthen the validity. The end result is not simply the data, presented in report format, but rather an analysis of these data and an interpretation of their meaning – in other words, a full evaluation needs to be conducted.

In most cases, it takes several years for a program to have a visible impact on its expected outcomes. For this reason, the standard practice of yearly reporting on results, starting from the first year of implementation, is not realistic. Depending on the program, the first few years of reporting are better to focus on implementation, with demonstration or results only beginning in year 3.

Competitive RBM stresses the accountability relationship between grantor and grantee. The information produced is not useful for program management because it narrowly focuses on a small set of key indicators, which are made as specific, measurable and unambiguous as possible (as well as “objectively verifiable”), and much effort is put into weighing the evidence for and against results achievement. Information is not produced in a timely manner, and is limited in scope, both of which limit their use for management purposes. Program management requires up-to-date

information with more emphasis on comprehensiveness than on rigour.

This does not mean that the program level cannot institute its own program-level RBM system, independent of the Competitive RBM relationship between it and its funder. But these RBM systems should be completely independent, relying on different types of information. Programs should not be accountable to demonstrate performance based on information generated through the program-level RBM system. If it is to be used for management purposes, there should be no incentives to bias the findings.

Improvement-Focused RBM

Improvement-focused RBM places greatest emphasis on encouraging program managers to manage. The accountability relationship is one based on learning and improvement, which reinforces program management. The benefit from RBM is had at the local level – it is intended for local level management, and therefore can only work in a decentralized environment where authorities to make decisions regarding program design and implementation have been delegated to the local level. The role of the centre is to create incentives for the local level to manage for results. Unlike Competitive RBM, Improvement-Focused RBM is formative in nature: performance information is used to improve programming in an ongoing manner.

Under Improvement-focused RBM, the central level holds programs accountable for learning and improvement: programs must show how they are using performance information to learn and improve upon their program delivery in the following cycle.

THE FOUR REASONS FOR POOR PERFORMANCE

REASON FOR POOR PERFORMANCE	APPROPRIATE RESPONSE
<p>1. <i>THERE IS NO PROBLEM WITH PROGRAM PERFORMANCE; THE FAULT IS WITH THE INDICATOR, WHICH DOES NOT ADEQUATELY CAPTURE THE RESULT INTENDED.</i></p>	<p>REVISE THE INDICATOR. THIS MAY OFTEN OCCUR WITH MORE COMPLEX AND ADAPTIVE PROGRAMS.</p>
<p>2. <i>THE INDICATOR IS APPROPRIATE, AND PROGRAM ACTIVITIES AND OUTPUTS WERE CARRIED OUT AS PLANNED, BUT EXTERNAL FACTORS BEYOND THE CONTROL OF PROGRAM MANAGERS PREVENTED THE DESIRED OUTCOMES FROM BEING BROUGHT ABOUT.</i></p>	<p>DEVELOP A PLAN OF ACTION TO MANAGE THESE EXTERNAL FACTORS IN FUTURE, OR IF THIS IS NOT POSSIBLE, REVISE OR TERMINATE THE PROGRAM. IT MAY BE THE CASE THAT THE EXTERNAL FACTOR THAT GOT IN THE WAY OF RESULTS ACHIEVEMENT WAS A ONE-TIME EVENT, UNLIKELY TO BE REPEATED. IN THIS CASE IT MAY BE REASONABLE TO MAKE NO CHANGES TO PROGRAMMING.</p>
<p>3. <i>THE PROGRAM WAS NOT IMPLEMENTED AS PLANNED. PLANNED ACTIVITIES AND OUTPUTS WERE NOT COMPLETED ON TIME, TO THE LEVEL OF QUALITY REQUIRED, OR OTHER CHANGES WERE MADE TO THE PROGRAM PROCESS DUE TO FACTORS INTERNAL TO THE PROGRAM.</i></p>	<p>INVESTIGATE THE CAUSES OF THESE UNEXPECTED VARIATIONS, FOR EXAMPLE THROUGH A PROCESS EVALUATION OR MORE INFORMAL METHODS. IT MAY BE THAT THE ORIGINAL EXPECTATIONS WERE UNREASONABLY HIGH, THAT THESE EXPECTATIONS WERE NOT FULLY UNDERSTOOD OR COMMUNICATED TO ALL INVOLVED, OR FOR SOME OTHER REASON. THE APPROPRIATE RESPONSE WILL DEPEND ON THE CAUSES FOR THESE VARIATIONS. IT MAY ALSO BE THAT, DURING PROGRAMME IMPLEMENTATION, THE NATURE OF THE PROGRAMMING CHANGED AS IT ADAPTED TO NEW EXTERNAL REALITIES. IN THIS CASE, IF PROGRAMMING IS SIGNIFICANTLY ALTERED FROM ITS ORIGINAL DESIGN SUCH THAT THE LOGIC MODEL NO LONGER ACCURATELY REPRESENTS IT, IT MAY BE IMPORTANT TO DEVELOP A NEW LOGIC MODEL AND NEW INDICATORS.</p>
<p>4. <i>THE LOGIC OF THE PROGRAM LOGIC MODEL DOES NOT HOLD: OUTPUTS DO NOT LEAD TO EXPECTED OUTCOMES.</i></p>	<p>DEVELOP A SOUND UNDERSTANDING OF WHY THE PROGRAMME LOGIC DOES NOT WORK AS EXPECTED (PERHAPS THROUGH AN EVALUATION), AND USE THIS NEW KNOWLEDGE TO IMPROVE YOUR PROGRAMMING. YOU MAY HAVE TO DESIGN THE PROGRAMME ANEW, INCLUDING DEVELOPMENT OF A NEW LOGIC MODEL</p>

Accountability for learning and improving means understanding the four reasons why negative outcome data may be arrived at, and responding to each in the appropriate way. Although it is often thought that poor outcome performance means the program itself is “bad”, this is not necessarily true. Data indicating you are not reaching desired levels of outcome achievement may be due to one of four reasons: the fault may lie with 1) the indicator used, 2)

external factors influencing results achievement, 3) the way the program was implemented, or 4) the program logic.

After each round of information generation and use, the program stands to improve incrementally in three areas: the implementation of the program, the theory of program change, and the performance measurement tools that link these two.

Accountability for learning and improvement means that managers are not accountable for demonstrating results, but rather for demonstrating that they are acting in accordance with the table above. In other words, the local level does not present performance reports with results information that “demonstrates” achievement, as per False RBM. Instead, it may present reports showing what it has learned from performance information, and how it reacted to this, or proposes to react, in adjusting its implementation or performance measurement methodology. The central level judges the program or unit based on how it reacts to performance information, not based on the performance information itself.

This type of accountability assumes that a program that is able to learn and improve may have more promise than a program that is able to demonstrate (or fabricate) results. Accountability for learning and improvement forces a change in culture from compliance to learning. It communicates a strong message from the top that RBM is about more than filling in boxes on a piece of paper that is quickly shelved: it is the method by which you grow and continuously improve your program.

Because of the nature of this RBM type, where performance information is used to inform corrective action on a program, timely and up-to-date information is of most value.

Unlike Competitive RBM, the quality of the evidence in favour of results achievement is of a lower priority to timeliness and comprehensiveness of information on results. Managers need information in real time to inform ongoing decisions. When a decision needs to be made, it will be made whether information is available or not, and in these cases information that has limited evidence supporting it is preferable to no information at all.

RBM conducted at the program level requires more than “objectively verifiable indicators”; a richer description is needed of the results achieved, both expected and unexpected, and the context surrounding them. The manager needs to understand what is happening on the ground, and a simple count of things or events provides too little information to be useful. With this added complexity of information, however, appropriate methods of analysis present a challenge, since program managers are not necessarily social science researchers. However, methods need not be too sophisticated. There are good methods already developed that capture the richness of detail needed in a timely manner and that are feasible in a development environment. An example of this is the Most Significant Change (MSC) technique, in which stories originating from beneficiaries, personnel or other stakeholders are collected and ranked to find the best stories on results being achieved. This technique includes a method for results aggregation, and so may also be suitable for Strategic RBM purposes.

Information on activities, outputs and short-term outcomes are often more salient to this RBM type than is information on medium-term or longer-term outcomes. For example, for a training program, immediate information on how well participants received the training and knowledge gained (or not gained) may be of more help in making ongoing pro-

gram adjustments than information that comes much later, with more tenuous links to program outputs, on the longer-term results of this training.

Improvement-Focused RBM presents an opportunity for more participatory development of results and indicators than may be possible with the other two RBM types. It exists primarily at the level of the program, and drawing in intended beneficiaries and other stakeholders may allow for a gradual coming together of perceptions about the program and its results, through iterative and participatory meetings on expected and actual program results.

Improvement-Focused RBM addresses some of the key reasons why RBM as it is practiced today is not used for program management. Program managers sometimes complain of out-of-date logframes bearing little resemblance to reality to which they are forced to adhere, and of onerous data collection requirements that appear to add little more knowledge of results being achieved than a simple visit to an implementation site might bring.

With Improvement-Focused RBM, these issues are alleviated: firstly, the initial logframe may only be a rough draft, since it is understood that it will improve over the course of the program; that is one of the side effects of RBM implementation. This reduces the upfront paperwork required and avoids the requirement to stick to an out-of-date concept of the program.

Secondly, while data collection may still be more onerous than simple observation, it will be beneficial to the program manager. Systematic data collection will serve not only for learning at the local level but also as a form of justification for action to the central level. In other words, in annual plans and performance reports, the program manager may present

performance information together with a summary of changes to program structure and direction, one as justification for the other.

Conclusion

Many organizations are unable to undertake the type of RBM that they claim to do. There are many granting agencies that require funded recipients to create a logframe prior to project implementation, and who require these recipients to “manage for results”, but simply do not have the bureaucratic flexibility to allow this to happen. In some cases, there may be good reasons for this, as the risks associated with it are simply too high. In these cases, however, there is simply no point in requiring a logframe, reports on “performance”, or any other product of (False) RBM at the level of the program. If the recipient is not allowed the flexibility to alter program design without central-level approval, there can be little real management for results at the local level. In this case, Strategic RBM would be of much greater value. If the central level has primary decision-making authority, it should aggregate results information to inform its own decision-making.

When implementing any RBM system, two important points should be kept in mind: firstly, do not confuse the tool with the user. Performance measurement is a tool useful for decision making; it is not the decision maker. A performance indicator, and any other information generated on results, is an input into a decision that must remain in the hands of a person.

Secondly, there are several tools in a manager's tool belt. Results information is important, but other things are important as well, such as information on process, cost, and stakeholder attitudes. All these need to form part of a manager's complex worldview in making decisions. Performance information is but one other aspect that can help a manager make better decisions.

CHAPTER FOUR

Evaluating Partnerships in the Not-for-Profit Sector

*Charles Lusthaus, Katherine Garven,
and Silvia Grandi¹⁸*

Because of the complexity of problems in our current world context, organizations of all types are increasingly recognizing a need to work together. Over the past decade, we have seen the growth of a wide assortment of organizational forms to tackle global challenges. These new forms are, in fact, constellations of organizations. Some organizations come together to formally create new entities and others do so less formally. Individuals and organizations in the field of international development are increasingly forging linkages with others in the public, not-for-profit and even for-profit sectors, to enhance their collective capacity in the hope that together they will better achieve their objectives.

A plethora of labels have been applied to these organizational groupings, including networks, consortiums, strategic alliances, coalitions, joint ventures, partnerships and inter-organizational relations. For the purpose of this chapter, we utilize the term “partnership” to identify these organiza-

¹⁸ This chapter is partially adapted from a paper written by Charles Lusthaus and Christine Milton-Feasby. (Lusthaus C & Milton-Feasby C (2006). The evaluation of inter-organizational relationships in the not-for-profit sector. Universalia. <http://www.universalia.com/site/files/ior-notforprofit.pdf>)

tional groupings. We have chosen to use the term partnership because it is the most commonly used term for these types of groupings, in particular in the field of international development. However, the reader should be aware of certain challenges in the use of this term.

The notion of partnership is common to many sectors and disciplines. Sociology, economics, political science, social psychology, and professional areas such as education, health, social work, and development studies include notions of partnership. Perhaps because of its rampant use, the term partnership has become fraught with ambiguity and added meaning. One finds in the field of international development that any inter-organizational relationship can be dubbed a partnership. Perhaps more regrettable is the extent to which the term has become value-laden. In the rhetoric of the international development field, a partnership has come to imply all things good. Equated with partnership are the virtues of equality, reciprocity, mutual benefit, and democracy. As Ostrower¹⁹ found, often the partnership is considered the end in itself, rather than a means to some end.

Another cause of ambiguity is the fact that in the corporate sector, partnership has become synonymous with contractual arrangements and is defined within legal parameters. In contrast, partnerships within the field of international development may or may not be formal arrangements. Public and not-for-profit agencies often enter into partnerships with or without few formal trappings. Public and not-for-profit agencies typically pay little attention to the legal framework set forth in private sector partnership law.

¹⁹ Ostrower, F. (Spring 2005) The reality underneath the buzz of partnerships. The potentials and pitfalls of partnering. *Stanford Social Innovation Review*, pp.34-41. (working paper)

In order to reduce the room for ambiguity and to narrow our focus, this chapter concentrates on organizational partnerships in the not-for-profit sector.²⁰

Over the past few years, we have been more and more involved in evaluating partnerships.²¹ As we did, we became more aware of the challenges related to these evaluations. We wondered about the questions that evaluators are asked to address and the basis of the judgments that they in turn render. At the same time, we began to ask ourselves about the very nature of partnerships: What are they? How do they form? Why do some develop and prosper while others flounder at start-up? What are the critical factors in their development and success? What do we know to date about this organizational phenomenon?

We reviewed the literature and now appreciate that while a great deal is written about partnerships, the evaluation of this organizational form is still an emerging topic. We realized that it would perhaps be useful to consolidate our experiences and put forward the ideas that have emerged through our personal experiences. This is the purpose of this chapter. First, we present some common characteristics of partnerships. Second, we outline some factors that we have noticed contribute to the successful formation and development of partnerships. Third, based on our observations, we present some implications for evaluating partnerships and discuss some common challenges that we have faced while evaluating partnerships and possible ways to address them.

²⁰ A significant issue in creating a better understanding of partnerships is the determination of whether the relationships within the partnership are of an organizational or an individual nature. This is not a trivial matter, as it affects many aspects of how a partnership works (e.g. in terms of governance structure) and how relevance and success are defined.

²¹ We have evaluated 20 different partnerships and consulted with IDRC as it did a major review of its own work on networks.

What is a Partnership?

According to our definition, a partnership is created when organizations come together collaboratively in order to address an issue of common concern. A new entity, often known as “the Partnership” is created with its own distinct governance structure, goals, and objectives. Despite the breadth of this definition and the diversity that it allows, partnerships share several characteristics. Concisely, partnerships are collectives of organizations that are voluntary, goal-oriented, complex, and flat in their authority structures. The relationship must benefit both the individual members and the partnership as a whole.

Partnerships are voluntary arrangements in that the member organizations come together of their own accord. It may well be true that the accomplishment of a specific objective obliges certain partners to join forces, as explained by Caplan²². However, this is not to say that the partners in such a case were forced to collaborate. The option of not undertaking an initiative always remains.

Partnerships are goal-driven collectives. Organizations come together in order to accomplish some specific objective. At the heart of any partnership is the appreciation of a compelling mission and the realization that none of the partners can achieve the mission alone. Indeed, it is further assumed that by collaborating, the members will achieve results that surpass the sum of the members’ efforts when acting independently. Synergy is therefore expected to flow from

²² Caplan, K. (2003) The purist’s partnership: Debunking the terminology of partnerships. *Partnership Matters*. BDP Practitioner Note Series. pp.31-36. <http://www.bpd-waterandsanitation.org>

the collaboration, and as it does, the objective becomes more achievable.

Partnerships are complex entities. At the same time that the member organizations are committed to the partnership's objective, they remain committed to the mission, goals and objectives that are unique to their own organization. Although individual organizations may align with the goals and objectives of the partnership, this is not always the case. Members therefore face the common challenge of having dual allegiances. Their purposes, structures, systems, and processes exist at the plane of the collective and at that of the component organizations. Further, the environmental contexts of the partnership are multiple. They include the context in which they function, plus those that influence each of the individual members. These environments may be the same, overlapping, or distinct.

Partnerships are hierarchically flat. Their flat governance structure reflects their collaborative origins. Members, who agree to share the costs of the collaboration, expect to share the responsibility of directing the activities of the collective. While such a structure signals respect for the individual members, it adds considerably to the complexity of directing and communicating throughout partnerships. It is often the case, for instance, that each member has veto power over the direction of the partnership.

Finally, partnerships must benefit the member organizations and the collective of the members as well as the partnership itself. When organizations agree to contribute their resources and expertise to a collaborative venture, they expect to be rewarded. Once again, the expected rewards are dual. Members expect to benefit from the accomplishment of the partnership objective and they expect that each organizational participant will benefit locally in the pursuit of its own

goals and objectives. Of interest here, we note that it is often unclear how individual members will assess these various benefits.

SUGGESTED FEATURES BY WHICH TO CLASSIFY PARTNERSHIPS

FEATURE	DIMENSIONS
SECTOR	▪ PRIVATE, PUBLIC
MOTIVATION	▪ PROFIT, NON-PROFIT
MEMBER INCLUSION	▪ NARROW, BROAD
DRIVERS BEHIND INCEPTION	▪ DONOR ▪ MEMBERS MARKET
ORGANIZATIONAL FORM	▪ SCALE (SIZE) ▪ HOMOGENEITY, HETEROGENEITY OF PARTNERS ▪ LOCAL, NATIONAL, INTERNATIONAL SPREAD
EXPERTISE	▪ PARTNERSHIP PROVIDES EXPERTISE FUNCTIONS OR COORDINATES EXPERTISE SUPPLIED BY PARTNER ORGANIZATIONS
RELATIONSHIP AMONG MEMBERS	▪ INDEPENDENCE (LOOSELY LINKED), DEPENDENCE (TIGHTLY LINKED), INTERDEPENDENT (MULTIPLE LINKS)
DEVELOPMENT STAGE	▪ BIRTH TO DEATH
STABILITY OF ALLIANCE	▪ TEMPORARY, PERMANENT
AUTHORITY	▪ DISPERSED, CENTRALIZED
STRUCTURAL ARRANGEMENTS	▪ SECRETARIAT, MECHANISMS FOR COORDINATION, REPORTING AND COMMUNICATION (TASK FORCES, COMMITTEES)

As a final comment, we emphasize that hopes run high when partnerships form. At the same time, so do the demands that are placed upon them and upon their member organizations. Due to the forces driving the creation of the partnership – the belief that complex issues can only be solved through the creation of a partnership – there is a great

expectation that partnerships create results quickly and demonstrate the benefit of collaboration over working individually.

Given the breadth of organizational relationships possible, the study of partnerships begs for one or more typologies. The following table contains a list of features that might be useful in sorting and classifying partnerships. This list is by no means exhaustive. It is meant to reflect the array of dimensions about which useful typologies might be created as a basis for the study of partnerships.

Learning from Experience: Some Observations

Our experience dealing with and evaluating organizational partnerships has resulted in several key observations, which we present in the following section.

Observation 1: Partnerships go through developmental phases similar to people and organizations.

Partnerships, just like people and organizations, go through stages of development. Understanding these stages can help with the successful planning and operating of partnerships. Expectations of the partnership should be realistically aligned to its stage of development. We have found that there are generally five stages of development: formation, growth, maturity, renewal, and decline.

THEORETICAL STAGES OF DEVELOPMENT OF PARTNERSHIPS IN THE NOT-FOR-PROFIT SECTOR

	MODE	LEADERSHIP ROLES	CLIMATE	DEVELOPMENTAL OBJECTIVES	PERFORMANCE OBJECTIVES
STAGE 1	FORMATION GETTING TOGETHER	CHAMPION	EXUBERANCE	<ul style="list-style-type: none"> ▪ TO ARTICULATE OBJECTIVE REQUIRING COLLABORATION. ▪ TO DETERMINE PARTNERSHIP'S NICHE AND POTENTIAL ORGANIZATIONAL MEMBERS. ▪ TO ENCOURAGE COLLABORATION. ▪ TO GET STARTED. 	TO START UP.
STAGE 2	GROWTH GETTING TO WORK	CULTIVATOR	PRODUCTION-ORIENTED	<ul style="list-style-type: none"> ▪ TO CLARIFY ROLES, RESPONSIBILITIES AND EXPECTATIONS. ▪ TO SET UP BASIC COORDINATION MECHANISMS. ▪ TO PRODUCE GOODS AND SERVICES. ▪ TO CREATE A STRUCTURE THAT FACILITATES ACTION. ▪ TO REFLECT ON THE BUSINESS MODEL. 	TO BEGIN SERVICE OR PROGRAM DELIVERY.
STAGE 3	MATURITY ORGANIZING OURSELVES	CONSOLIDATOR	RESULTS-ORIENTED	<ul style="list-style-type: none"> ▪ TO DEFINE THE BUSINESS MODEL. ▪ TO INSTITUTIONALIZE MECHANICS FOR WORK PLANNING, SHARED DECISION-MAKING AND COMMUNICATION. ▪ TO ESTABLISH FORMAL EVALUATION AND MONITORING SYSTEMS. ▪ TO TEST EFFECTIVENESS AND IMPACT. 	TO SHOW THAT OUTCOMES CAN BE ACHIEVED.
STAGE 4	RENEWAL RE-COMMITTING OR REFOCUSING	CHANGE AGENT	REINVIGORATED	<ul style="list-style-type: none"> ▪ TO RECOGNIZE SIGNS OF A PARTNERSHIP IN TROUBLE. ▪ TO ENCOURAGE SETTLING OF DISPUTES OR REVITALIZE PARTNERSHIP AROUND A FRESH PURPOSE. 	TO INCREASE REACH, INTRODUCE NEW SERVICES OR PROGRAMS.
STAGE 5	DECLINE COMING APART	PHILOSOPHER	DESPAIR & ACCEPTANCE	<ul style="list-style-type: none"> ▪ TO CREATE A NEW PARTNERSHIP (PURPOSE, PARTNERS ETC.) ON FOUNDATION OF OLD RELATIONSHIPS. ▪ TO ORCHESTRATE ITS DISSOLUTION, SO THAT GOOD RELATIONS ARE MAINTAINED AMONG PARTNERS. 	TO TERMINATE SERVICE OR PROGRAM.

The formation stage focuses on gathering members and defining the partnership's goals and objectives. The general atmosphere is most often very positive and optimistic with leaders championing the cause. The second stage, which is one of growth, is when the partnership begins to operationalize. Roles and responsibilities are clarified and mecha-

nisms for cooperation and collaboration are established. This is the stage where the partnership begins to deliver initial outputs towards the partnership's goals and objectives. The maturity stage is when significant outputs are expected from the partnership. The partnership then typically becomes results-oriented and often establishes formal monitoring and evaluation systems to ensure high performance. This is a crucial stage in the development of the partnership in that operational mechanisms become formalized. After a period of maturing, the partnership typically enters into a stage of renewal where it refocuses its activities towards its initial goals and objectives. During this stage, there is often a sense of reinvigoration among members where they often introduce new activities and approaches. The final stage is known as the decline stage where the partnership eventually dissolves. Members often try to maintain good relations with one another and even think of ways to create new partnerships based on the lessons learned from the previous partnership.

Observation 2: Partnerships form to serve the greater good, but must work together on focused and targeted objectives if they are to be successful.

Partnerships form around greater missions. They do so because the partners realize that alone they are unable to manage the scope or type of activities necessary to accomplish a prized objective.

While this sounds easy, it is not. Many partnerships have an extremely difficult time focusing their work. Partnerships are created because the partners believe that in doing so they can address big issues. However, once the members begin to address issues, they often realize that the partnership –

while better able to deal with the issue – is still limited. Persisting to strive towards broad, long-term goals cause the partnership to become bogged down and lose momentum. Paradoxically, partnerships are better able to engage in their initiatives over the long-term when they set and work towards more narrow, immediate goals. Successful partnerships are pragmatic in this regard. They match expectations with resources. They make tactical and focused adjustments throughout their existence.

As one example, Ugandan women's groups got together to address the lack of access to education for young girls in their country. No one group was able to deal with this issue alone. Moreover, they felt the government was also incapable of taking on this issue. Similarly, a partnership was formed of environmental organizations to significantly increase the number of assessments of species around the world so that they could better identify species in danger of extinction. As a final example, a group of organizations dedicated to promoting the rural poor's ability to secure access to land formed a coalition with this objective. They realized that the success of their objective hinged on changing international and national laws and enforcement mechanisms – a step that none of them could hope to accomplish alone.

Observation 3: Since most not for profit partnerships are value-based, trust is the major underlying glue that holds most partnerships together.

Trust is necessary for the establishment and maintenance of partnerships. Although trust is an important issue in organizations in general, it is particularly important in the formation and operationalization of partnerships. This is

because members of a partnership come together voluntarily to collaborate in addressing an issue of common concern. The collaborative nature of a partnership results in the diffusion of responsibility and accountability among members. Therefore, the success of one member depends on the performance of the other. Since membership is of a voluntary nature, it is difficult to enforce positive collaboration and good performance from members, therefore creating the need for trust among members.

Trust is of particular importance in the formation stage of a partnership, since trust must be present before an organization agrees to enter into a partnership. However, trust remains an important issue throughout the partnership's evolution. During the growth of a partnership, trust builds, sometimes comes to be taken for granted, is often tested, and builds afresh. No matter how the issue of trust may evolve throughout the life cycle of a partnership, it is always a necessary factor for the successful performance of a partnership.

As a corollary, without open and candid communication among partners, there will not be trust. It is our experience that trust is built through consistent and predictable communications. Since open communication is the source of trust, successful partnerships build and utilize effective communication systems. This is especially true in complex partnerships involving diverse members. In these cases, the partners may not be natural allies. In addition to bringing their strengths to the partnership, they also bring preconceptions, protectiveness of turf and suspicion. Consequently, trust can be strained and the relations among members can be precarious.

It is important to point out that in the not-for-profit world many factors work against creating these trusting relationships. First, partnership members often compete for funding. Second, most partnerships are partly supported by large government bureaucracies that in turn impose a host of bureaucratic requirements. Bureaucracy does not facilitate trust. What appears to assist members in overcoming these roadblocks is the creation and fortification of interpersonal ties among key individuals. When the individuals who are most deeply involved in the partnership forge personal connections, system trust gradually evolves.

Observation 4: The successful formation of partnerships depends in large part on the vision, commitment, drive, and interpersonal sophistication of individuals who champion and lead the venture.

This observation concerns leadership and is integral to the successful formation of partnerships. Champions play a critical role in forging these relationships. Individual champions are often the starting-point for the formation of partnerships. Typically, it is a select group of men and women who appreciate early on the potential of partnerships to make possible the accomplishment of some broad mission. They anticipate the benefits of such partnerships. Despite being aware of the difficulties that such relations promise as well, they take on early leadership roles that are time-consuming and potentially injurious to their organizational careers.

In all of the partnerships that we evaluated or re-viewed, champions were a critical factor at start-up. An interesting question is whether they stay beyond start-up. In our cases, we found a variety of practices. Ultimately, the partnership needed at least one champion to stay over the first three stages of the evolution of the partnership. The champion needed energy to persevere. Of interest is that few had the managerial qualities to enable the network to move into a more mature stage. Hence, in almost all instances problems arose in Stage III.

Observation 5: Successful partnerships invite the right partners at the beginning, embrace new partners if need be, and allow partners to exit when appropriate.

It is critical that the right partners are involved in the partnership. They should do so from the start. By right we mean organizations that embrace the mission of the collective and have a significant contribution, e.g. resources, knowledge, legitimacy, commitment, expertise etc. to make towards it. Further, it is important to identify from the start not only organizations whose contributions are required in the early phase of the relationship, but also organizations whose contributions may become important in later stages.

In the event that necessary partners are forgotten, they must be identified and invited in later. To do so, the partnership needs to establish mechanisms to review the constituency of the partners over the life of the partnership. These mechanisms enable the members to discuss their contribution and willingness to participate. To do so, the partnership needs to nourish a climate of openness so that the partners feel comfortable both assessing others and being assessed by

them. We appreciate that it is difficult to bring new partners into the partnership. The process of cultivating trust between partners is laborious and time consuming. Therefore, it becomes increasingly more awkward to introduce newcomers. Typically, we find that new partners already have linkages or a history of collaboration with incumbent members. In that way, some trust has been institutionalized.

Throughout our work we found that membership matters. Indeed, the members who started a partnership might not be the right members at a later stage. For example, when we explored a network of health agencies, we found that their network was missing some of the most important players in their region. Without these organizations joining the partnership, the objectives of the partnership could not be met. It took the evaluation to say this and lead to the growth of the partnership.

Finally, it is vital that partners be allowed to leave gracefully so that future collaborations are not jeopardized. Review mechanisms are necessary here again to enable partners to assess their ongoing contribution. At the same time, such formal mechanisms provide the members with a platform from which to announce their intentions to withdraw in the event that they no longer feel valuable to the collective.

Observation 6: Successful partnerships adopt a business model that aligns resources with their goals.

Often, the work of partnerships is initially financed through short-term, project targeted funding. Successful partnerships, however, take advantage of this early financing to

reflect on longer-term funding needs as well as the disbursement of these funds. The partnership business model needs to emerge within a year or two of start up even though it is difficult to resolve because all members have similar resource issues.

In our evaluations of partnerships, we have witnessed real neglect of the business model. Partnerships in the not-for-profit sector often start because of financial opportunity. In other words, a donor puts money on the table. However, the seed money is often spent within a couple of years. Continuation of the partnership depends on the willingness and the ability of the members to create a viable financial system to support the work of the partnership. Few partnerships, that we have seen, engage in serious business planning. Our experience indicates that they often limp along - waiting for the next donor. Sustainability requires otherwise. It is our observation that partnerships in the not-for-profit sector need to devote more time and effort to the definition and development of their business model.

The business model links resources to activities. The three basic models are donor driven, membership driven or product and services driven (market driven). In donor driven arrangements, a funding agency or agencies supply the financing. This agency may or may not participate in the work of the partnership. Regardless of whether the donor takes part in the operations or not, it will have considerable influence over the partnership. Alternatively, the members may choose to support the partnership by drawing on their own sources of income. Discussion will revolve around the pooling equation. This choice may strain the partners' resources, but leaves the power in their hands. The fiscal

burden on the partners and their capacity for self-determination vary in hybrid models (combining donor-driven and membership-driven models), depending on the degree to which the partnership is internally or externally funded.

For example, in reviewing a partnership for school improvement, we found that schools and school districts were constantly redefining inclusive schooling, which means the accommodation of “all” handicapped and disabled children in public schools. Not unpredictably, “all” meant different things to different people. This led to a questioning of common values and principles. Flexibility and perseverance on the part of the members resulted in an acceptable refinement of guiding principles. These principles led to concrete training proposals that all members saw as a step in the right direction. The members were able to get behind the partnerships and support the training. “It is our training,” one member remarked. As a result, the partnership remained intact.

Finally, the partnership can choose to sell its expertise to outsiders. Hence, the financing decision is both a control and a resource decision. The partners have to decide whether to increase their resource base by seeking outside funds or to rely on the resources they bring collectively to the partnership. If they choose to go outside for funds, they risk becoming dependent on the donor. As a result, the partnership is vulnerable to collapse should funding be withdrawn. Perhaps more importantly, the partnership may undergo a shift in direction under pressure from the donor to pursue the donor’s agenda rather than the partners’ original purpose. Alternatively, a partnership that relies on its collective resources may enjoy fewer resources than it might

otherwise. However, drawing from many different sources of income, the collective is less vulnerable to the loss of donor support and therefore may feel free to be true to its original purpose. Ultimately, in making their choice of business model, the partners have to be sensitive to the choice they are making between wealth and control. Their choice needs to be consistent with their collective values as well as recognize their need for funds.

Observation 7: Successful partnerships are able to provide products and services quickly. They need to show that their costs are worthwhile.

The operational and administrative costs of partnerships are known to be very high. High operational costs are due to the challenges in addressing complex issues while high administrative costs are due to the challenges in organizing multiple members. These costs, however, are typically justified through the belief that important issues of common concern can be better addressed through the creation of a partnership. In order for the partnership to obtain or to continue receiving credibility, it must demonstrate that this justification is in fact true.

Partnerships must demonstrate their worth as early as possible. Partnerships that fail to do so die! Creating feedback loops and opportunities for members to articulate their expectations enabled one of the partnerships that we evaluated to promote discussion and disclosure. In partnerships that lacked the mechanisms for open dialogue, we found that unclear expectations led to unnecessary conflict and in some instances the withdrawal of members.

Therefore, successful partnerships ensure that they deliver initial outputs quickly and maintain a good balance between long-term goals and “quick wins”. Providing credibility and justifying the high costs associated with the partnership is essential in ensuring membership participation and continued funding.

Observation 8: Partnerships are more likely to succeed when the partners take charge of the partnership and demonstrate ownership rather than giving ownership to paid managers.

While the vision of persuasive leaders and the allure of a compelling mission motivate partners to come together, the glue that keeps them together is a belief in the value of the partnership itself. Partners must experience ownership. When partners experience ownership, they believe that the collective belongs to them and not the reverse. They believe that they have the right to steer it and are not subservient to it. They believe that they are entitled to a fair share of the benefits that accrue and are not supplicants when it comes to allocating gains. They donate the promised resources without suspicion that others are benefiting at their expense. When they experience ownership, members are more willing to participate in the strategic planning and administration of the superstructure.

As a corollary, it is more difficult to create ownership in grossly heterogeneous partnerships. These are more fragile arrangements. Members in partnerships with a widely eclectic membership contain disparate value systems, goals, work styles, and speak from unique points of reference. As a result, the individual partners may fail to identify closely with the collective.

In contrast, where differences exist but are not profound and where practical considerations support moving toward shared objectives, partnerships can be sustained. This is particularly true where there are urgent, practical considerations that entice the partners to continue moving toward shared objectives.

Observation 9: In successful partnerships, members adopt the institutional arrangements and administrative systems that support the partnership.

As partnerships continue to operate and grow, elaboration of structure increasingly concerns the partners. In the early stages of development, the partners adopt the basic rules and procedures necessary to initiate the activity of the partnership. The issue relatively early is how to ensure the smooth operation of the partnership and thereby its ongoing activities. In subsequent stages, the partners focus on both ensuring the delivery of valued services and adjusting its institutions and systems to support changes in direction and activity. The basic issue is how to manage the partnership. Do you pay for a Secretariat? Should members take on managerial responsibilities? What is the balance?

Control is a prickly, but vital consideration. It is contentious in that the partnership members have authority over a secretariat but probably have little authority over representatives of the partner organizations or even over the partner organizations themselves. The partnership, however, does have responsibilities to members and vice versa. In fact, there are multiple levels of responsibilities within partnerships. Therefore, members of successful partnerships debate alternative management structures and choose the one that

reflects their collective values (be it preferring democracy or expediency and efficiency). They then imbue these roles with sufficient authority to provide direction and resolve conflicts. Alternative governance structures include the establishment of a formal Secretariat, a formal Steering Committee, rotating leadership by the partnership, or the appointment or emergence of a lead partner. Each choice has implications for the efficiency of the partnership and the distribution of power within it. As such, partners tend not to make governance decisions lightly.

Observation 10: Successful partnerships are mindful of the dual allegiances of the individuals within the system and take steps to alleviate the potential for role conflict.

While key individuals serve both the partnership and their own organization from the start, the impact upon these individuals and the system is especially felt in mature partnerships. Therefore, we most associate the problems of dual loyalties to Stage III. While in the early stages, key leaders speak for their own organization and the ideal of the partnership, it is as operations take hold that they are increasingly involved in the work of the partnership. Not only do they experience increasing demands on their time and energy, but they may be torn between their responsibility to represent the partnership and their organization.

Successful partnerships are both aware of the role conflict implicit in the multiple allegiances common to partnerships, and they install mechanisms to bring these conflicts to light and resolution. In these collaborative arrangements, partners are both governor and operator. Accordingly, individuals in the collective often wear two hats. At the most basic level,

individuals in the partner organizations have a role and specific responsibilities to the organization to which they adhere. At the same time, each organization can be said to have a responsibility toward its employees, volunteers, associates etc. At the next level, individuals in the partnership who represent the partner organizations (e.g. members of the Secretariat) have a role and responsibilities to the partnership as well as to the sponsoring organization from which they come. Here again it is imagined that the relationship is reciprocal to the extent that the partnership (e.g. the Secretariat) and partner organizations have a responsibility to the individual representatives. At the next level, the partnership can be perceived to be responsible to and having a role to play *vis-à-vis* the partner organizations. At the highest level, the partner organizations have a role to play and responsibilities within the collective. They are effectively responsible to one another.

This notion of responsibility to one another is the ideal. At any point, of course, these multiple allegiances may cause conflict. What is good for the partnership may not be good for a particular partner. What is the individual to do? Should he or she sacrifice the needs of his organization for the good of the whole or some partners? Should he or she favour the employer? Should a compromise be sought, and if so how? Successful partnerships anticipate, discuss, and work through the push-me-pull-you situations endemic in these systems. They attempt to do so through the strategic design of their coordination and communication systems. Nevertheless, they succeed, if they succeed, through the operation of trust.

Observation 11: Successful partnerships are mindful of the transaction costs that plague such systems and take steps to anticipate and budget for them.

As the previous observation, this is another cautionary observation, focusing on cost recognition rather than cost control. Like with the preceding observation, transaction costs are particularly felt by Stage III. In previous stages, members may profess sensitivity to the increased administrative costs of networks. However, the impact of these escalating and often hidden transaction costs is not truly felt until partnerships reach maturity. As the structure becomes more elaborate and the systems demand greater input from partner representatives, the partners and their representatives feel the strain of limited time and resources to carry the administrative load. In subsequent stages, these costs continue to be material and the topic of debate.

Partnerships require enormous inputs from the partners and are particularly unwieldy organizational forms. As a result, their coordination costs are onerous. Their processing and reporting costs are high due to the multiple layers of interested parties involved coupled with the lack of firm authority and direct reporting lines. The logistic demands of partnerships mount with the geographic spread of partners. Given the complexity, there is the potential for a number of costly dysfunctional outcomes such as resource hoarding, the gravitation towards risk-averse agendas and free riders. Successful partnerships are aware of these risks from the outset and build mechanisms to anticipate and counter them.

A second problem is the relative invisibility of these transaction costs in comparison to the direct costs of programs and

services. Partnerships redirect employees' (and often key employees) time from the partner organization to the partnership itself. The costs of their time may be neither recognized nor accounted for. However, this time is lost to the partner organization. At some point, certain partners may have insufficient slack in terms of their staffing to continue to bear this burden, with or without rewards from the collective.

Transaction costs can overburden partnerships and contribute to their decline, so to succeed, partnerships must pay these costs more than lip service. Partnerships need to recognize the potential for growth of transaction costs at the outset, plan for them by preparing systems to track them and agreeing to allocate funds to defray them before they accumulate. The creation and use of feedback systems (monitoring and evaluation) by the partnership inspire trust amongst the partners, which again helps sustain the partnership.

Observation 12: Partnerships are bred in the promise of synergy and learning. Thus, when synergy fails to emerge, partnerships are in peril.

Like the observation before, the final observation is an observation of performance gone wrong. Successful partnerships help people and organizations learn, and through this learning produce synergy. This means that their output exceeds the sum of what the members can achieve independently. Members of partnerships that yield synergistic results stay together and continue to work together. Conversely, partnerships that fail to yield the benefits of synergy, dissolve. Feedback loops such as monitoring and

evaluation systems help in this. So do a wide assortment of formal and informal communication tools.

The promise of synergy is the operational rationale for partnerships. It is this promise that attracts partners to collaborate. It is this promise that encourages them to maintain the collaboration despite the enormous transaction costs inherent in such complex organizations. Because of the substantial and accumulating system costs, the output of partnerships must be substantial as well. When partners join, they have some inkling of the expense of union. Each of them knows what resources they can offer. In addition, each of them understands what they can accomplish on their own. When they agree to collaborate, the partners are gambling on the benefits of cooperation. Their wager is that the results they produce together will surpass the sum of their independent efforts. The cost of this wager is the cost in excess of the direct costs of the members' activities. These excess costs are what we refer to earlier as transaction costs. Often they are hidden. They are difficult to quantify. Finally, these costs rise rapidly over the life of the partnership. Thus, where there is no synergy or the benefits are only slight, the costs of the collaboration may swamp the benefits of a partnership. For this reason, partnerships that disappoint in the production of synergy will fold.

Implications for Evaluating Partnerships

In recent years, there has been an increased interest in evaluating partnerships. However, the specific characteristics of partnerships that we have highlighted affect the way evaluation methodologies, approaches and criteria can be adapted to them.

Furthermore, international development is changing, and is expected to change more in the future. The following table provides a short history of how the field of evaluation in international development is turning its attention towards the evaluation of partnerships.

*THE CHANGING LANDSCAPE OF INTERNATIONAL EVALUATION*²³

ORIENTATION	PAST	PRESENT	FUTURE
<i>FOCUS</i>	MONEY	POLICY	INSTITUTION
<i>UNIT OF ACCOUNT</i>	PROJECT	COUNTRY	GLOBAL
<i>STRUCTURE</i>	FRAGMENTED	COORDINATED	INTEGRATED
<i>INSTRUMENTS</i>	PROJECTS	PROGRAMS	PARTNERSHIPS
<i>RESULTS REQUIRED</i>	PROJECT OUTPUTS AND MAYBE OUTCOMES	PROGRAM OUTCOMES AND POSSIBLY IMPACT	CREATING GLOBAL CHANGE INSTITUTIONS AND QUALITY ISSUES

The evaluation of partnerships has become important during the last decade as there are more partnerships and they are being touted as the panacea for solving many problems. There is an evident symbiotic relationship between belonging to partnerships and the worth of individual organizations and it is becoming increasingly apparent that the embeddedness of organizations in partnerships might have direct implications for members' value. In other words, organizational relationships—through partnerships—might be a valuable resource for an organization. Thus, partnership members and other stakeholders are calling upon evalua-

²³ Adapted from Stern, E. (2004). "Evaluating Partnerships" from Eds. Liebenenthal et al. Evaluation & Development: The Partnership Dimension. World Bank Series on Evaluation and Development. Vol. 6.

tions to shed some light on the performance of partnerships. Funding agencies also increasingly want to understand the “full value” of partnerships. The evaluation of partnerships has become important for funding agencies and others who share an interest in knowing what the overall value of the partnership really is. However assessing all “value added” has proven to be very difficult to capture— and few have tried. Doing so however, is crucial!

For some time now, we have been living in an “age of evaluation”²⁴ or an “auditing society”²⁵ which increasingly requires the assessment of results. This reality explains the increasing demand from internal and external stakeholders for useful approaches and indicators to assess the performance of partnerships.

Results are a central theme in all types of evaluation work, but one of the big shifts in recent times is the recognition that “organizations matter” and that partnerships as an organizational form matter more and more. This new evaluation perspective reflects the increasing need for the evaluation community to develop concepts, tools frameworks, approaches and so forth to assess the performance of partnerships. Thus, while in the past evaluations borrowed their methodologies primarily from psychology and sociology in the partnership area they are drawing from a variety of fields, including the management sciences, transaction cost economics, resource dependence, or neo- institutional theory, in particular to meet its needs in evaluating partnerships. Progress is being made, however significant challenges remain. We have identified five such challenges.

²⁴ Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation Evaluation*. Thousand Oaks, Calif.: Sage.

²⁵ Powell, W. (1990). Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*. 12: 295-336.

Defining the Effectiveness of Partnerships

At the heart of every organizational form is a purpose or a goal. Organizations are set up to serve a reason and are ultimately judged on the extent to which they contribute or meet their objective. The same is true with respect to partnerships. They are set up to fulfill one or more objectives. Normally, the objective(s) is identified in a way in which none of the partners can reach the objective(s) without working together. Thus assessing the extent to which a partnership is moving toward its objective(s) is a direct approach to assessing the effectiveness of a partnership.

Effectiveness is often the focus of the partnership. Assessing effectiveness means answering the question “Did the partnership achieve what it was formed to achieve?” However, this is not straightforward. The plurality of stakeholders involved in partnerships often translates into a plurality of expectations and views on what should be a partnership’s key objectives. The reconciliation of these expectations is crucial to allow partnerships to be functional. However, this type of process results is rarely assessed. Members of partnerships are constantly assessing the benefits of the partnership, but they are also often assessing the benefits to their own organizations. Successful partnerships not only have to meet the expectations of the partnership, but they also need to meet the expectations of their individual members. It is the dual benefit, which needs to be understood and described when one assesses partnerships. Assessing one without the other misses important aspects of the partnerships.

In addition, because of the multi-stakeholder nature of partnerships and the variety of relations and reciprocal influences between the members and the partnership, it is difficult to determine whether, and to what extent, the partnership has contributed, directly or indirectly through its members, to the achievement of results. It is challenging to distinguish between the partnership's contributions and the contribution of members.

Members also expect from partnerships a systemic effect. They expect partnerships to be able to achieve something more than the sum of what individual members would be able to achieve individually. Measuring this systemic effect should be an important focus for the evaluator; however, there is no clear formula for doing so. Some partnership participants argue that partnership effectiveness is the sum total of the partnership work plus that of individual members. The argument is that when considering the partnership not only should the value added of the partnership be assessed but also the value that individual organizations contribute. The partnership challenge is to clarify "what defines effectiveness success"? The value added brought by the partnership or the totality of individual organizational results and the partnership. Figuring out which to evaluate is a challenge.

Determining the Right Basis to Evaluate "Success"

As mentioned previously, organizations that become members of partnerships face dual allegiances. Even if an organization has an interest in participating in a partnership, it does not mean that the organization shares all of the same

values and viewpoints of those that the partnership represents. Organizations may have their own particular motives for joining a partnership, which may not be completely aligned with the goals and objectives of the partnership. Therefore, when evaluators attempt to rate the level of “success” of the partnership, they need to be conscious of what judgment basis they are using to evaluate it. Should the criteria of stakeholder satisfaction be used even if the organizations (the stakeholders) are happy with the results (for instance, if the organizations received positive public approval), but the results have not necessarily met the partnership’s goals and objectives? Or should the level of success of a partnership be based primarily on whether or not the problem or issue has been solved, or at least to what extent it has been solved? We try to use a mix of both approaches. However, we are constantly faced with the issue of judgment basis when evaluating success, particularly in the case of partnerships.

Clarifying the Performance Framework

In general, evaluation standards have argued that there are at least four key concepts that make up a complete evaluation. They involve the evaluation of effectiveness, efficiency, relevance, and sustainability. We have presented above the issues related to evaluating partnerships’ effectiveness. In addition, the other criteria pose some challenges to partnership evaluation, as well as the selection of an overall evaluation framework that can be useful for partnerships. Evaluation research exploring these issues, however, is in its

early stages. Work by Creech²⁶ explored the concepts and techniques used in partnership evaluation. She found little consistency. Similarly a review of the literature by Gandori²⁷ indicates that partnership performance, if measured at all, has to date been captured at the organizational/structural level. In general, research exploring performance frameworks for partnerships has found mostly concepts related to structures and processes that contribute to the overall success of the partnership. These structures and process we would call capacities or management practices. For example, in our own assessment of a multi NGO partnership, we identified 13 criteria that could be used as a framework to assess the performance of a partnership. Of note is that structural and process concepts dominate the list.

On the more performance-oriented side, Creech found that partnership evaluation, besides effectiveness is often asked to assess efficiency, ongoing relevance, and sustainability. Efficiency is a bit tricky, as we know that in general, the coordination of partners often lead to high transaction costs. One of the central measurement issues for evaluations is to calculate such costs. However, what costs ought we measure? Some partnerships costs are inputs into other important processes that could very well be objectives of members. For example, North-South linkage partnerships are sometimes set up to build the capacity of the southern partner at costs to the northern partner. However, southern capacity building is often seen as outcomes required for even larger processes. Is success defined in relation to the larger processes or the partnership itself? The distinction between

²⁶ Creech, H & Ramji, A. (2004), *Knowledge Networks: Guidelines for Assessment*, IISD and Creech, H. and Willard, T. (2006). " Strategic Intentions: Managing knowledge networks for sustainable development". *Network Digest*.

²⁷ Grandori, A. (1998) Editorial: Back to the future of organization theory. *Organization Studies*, 19(4): i-xii.

antecedents and outcomes of some partnership costs may be misleading or not possible to track.

Criteria used to assess the Multi NGO Partnership

1. Partnership members share common principles
2. Clarity of structural arrangements
3. Processing of work is collaborative
4. Appropriate oversight system
5. Quality of communication networks
6. Benefits to members from the partnership
7. Partnership synergy
8. Distinct partnership identity
9. Transparency
10. Trust between members
11. Institutional leadership
12. Partnership alignment with its member organizations
13. Management model

Ongoing relevance presents another framework challenge. Clearly partnerships need to be relevant; but to whom? Relevance has various layers of complexity in partnerships. First, when an evaluator tries to assess relevance to the partnership's stakeholders, things can become complicated. Traditional definitions of relevance assume stakeholders are distinguished groups with clearly distinguished priorities. This assumption is not valid for partnerships, where in many cases the members are also the beneficiaries, the target groups and in some cases the donors and where the boundaries between what is inside and what is outside the partnerships are often difficult to draw. Moreover, when the members are organizations, each has its own sets of priorities, and is often represented by individuals, who also have their own personal agendas to satisfy.

Given the diversity and complexity of the membership of many partnerships, relevance often varies from one member to another. Given the voluntary nature of partnerships, being relevant to members is often seen as crucial for a partnership's survival. If members do not think that the partnership they belong to is being relevant to them, they will simply withdraw from it. Thus is ongoing relevance of function of membership continuity—or does it relates to the substance of the work. What should the concept denote?

Finally, sustainability, like other typical framework issues, plays slightly differently in a partnership. When we assess the sustainability of a partnership we are mainly interested in two dimensions: the ability of the partnership to secure over time the resources needed to achieve its objectives; and the ability of the partnership to maintain the internal and the external relationships needed to exist and function (this includes having the right members, securing their participation, maintaining a niche among complementary and concurrent players in the context). However, the question of sustainability when looking at partnerships is somewhat unique in that the purpose of the creation of a partnership is to address a particular issue or problem. Once that issue or problem has been addressed, there may no longer be a need for the partnership to exist and its sustainability is therefore no longer an issue. Unlike organizations, which are not time bound, the idea of the sustainability of partnerships is not well developed.

High Levels of Participatory Involvement but Lack of Clarity with Respect to Ownership

We identified the costs of coordination that is associated with efficiency as one of the real problems for a performing partnership. However, the participatory nature of partnerships presents a challenge not only to efficiency but also to the methodology of evaluation. In evaluating partnerships, it is important to ensure that the evaluation is of a participatory nature. This is because of the mix of voluntary participation, complex sets of expectations among members, non-hierarchical governance structures, and spread-out accountability systems within partnerships. Participatory M&E approaches are more suitable to identifying achievements comprehensively and integrating different perspectives, clarifying and mediating different expectations, fostering shared thinking on the partnership performance, building internal capacity, and circulating an understanding of M&E among partnership actors. Participatory approaches to monitoring and evaluation also push partnerships to think about the constellation of stakeholders that revolves inside and around them and to clarify their relationships and mutual influences.

Participatory approaches also play into the value systems often associated with the typical voluntary sector partnership we have evaluated. Members want to feel that they are intimately involved in the partnership. Motivation is driven by participation and vice versa. For many their participation represents their conception of ownership. Ownership is thus assumed by the energy put into participation.

However, evaluations are part of the decision making process. Evaluations ask about merit or worth. If worth is questioned who is responsible for making ultimate decisions about closing or growing? We have found that one of challenges in the evaluation of partnerships is who is ultimately the owners—the responsible parties. We experienced the energy of members who participate in partnerships, but these members are not often the leadership of their organization; and when hard resource decisions need making it is Senior Organizational Members who make them (often not those engaged).

As stated above, partnerships are typically formed to address an issue of common concern. In the not-for-profit sector this common concern is often in the form of a “global public good” such as security and safety, environmental issues, human rights, etc. However, the voluntary and participatory nature of partnerships has a tendency to diffuse accountability among the partnership members. Therefore, when the good is of a public nature and accountability is diffused, we ask the question: Who owns the results of the partnership? When everyone holds a stake in the outcomes and yet no one is particularly responsible, ownership can become vague and undefined. This is an evaluation challenges in understanding the dimensions of partnership ownership.

The Methodological Challenges

Partnerships are difficult to evaluate due to their highly complex nature. The attribution of results is difficult in multiple, inter-connected, and ever-changing contexts. The relationships of actors within and outside of the partnership and the multiple perspectives of these actors add another

level of complexity. As mentioned previously, there has been little research to date to create acceptable frameworks or models to help evaluators work in such environments. Similarly, most evaluations have followed the linear logic model approaches that use case studies and multiple data sources. This approach has dominated much of the international evaluation work. However, given the complexity of partnership arrangements, new methodological approaches are needed.

Morgan²⁸ and Tascherau and Bolger²⁹ argue that complexity theory offers some insights and practical tools to try to understand some of the “messiness” of partnerships. In addition, Wilson-Grau³⁰ stresses the fact that “increasing application of complexity science to the challenges of social change organizations offers important insights” to understand and assess partnerships.

Methodologies for assessing partnerships should be able to take “full account of the messy, multi-level, and multi-directional causality of the process and environment of implementing, monitoring and evaluating its strategic plans”. Westley, Zimmerman and Patton, in their book *Getting to maybe: how the world is changed* argue that “...to know step by step, in advance, how the goals will be attained [is] an approach doomed to failure in the complex and rapidly changing world in which social innovators

²⁸ Morgan, Peter. (2005). *The Idea and Practice of Systems Thinking and their Relevance for Capacity Development* and Morgan, Peter. (2006). *The Concept of Capacity* (Draft Version: Study on Capacity, Change and Performance)

²⁹ Taschereau S., Bolger J.(2007) , *Networks and capacity*, Discussion paper No 58C February 2007, ECDPM

³⁰ Wilson Grau R. (2008), *Complexity and International Social Change Networks*. In Sheers G. et al. (2008), *Global Partnership for the Prevention of Armed Conflict. Assessing Progress on the Road to Peace Planning, Monitoring and Evaluating Conflict Prevention and Peacebuilding Activities*, Issue Paper 5, May 2008

attempt to work. In highly emergent complex environments, such prior specification is neither possible nor desirable because it constrains openness and adaptability.”³¹

For this reason approaches and tools that have been inspired by complexity theory are particularly well suited for partnerships – e.g. broad goals, a 'good enough' vision and minimum specifications³² allow managers to work within a coherent and strategic framework, including prospects for building capacity while leaving space for flexibility and adjustment to diverse contexts and realities in which members operate.

In this context, Wilson-Grau argues that partnerships should invest in monitoring and evaluation as an ongoing, formative evaluation mode. This type of approach, known as “developmental evaluation”³³ would focus on continuously collecting and assessing data in order to tell what works and what does not and to be able to adapt accordingly. This approach is more of an iterative performance management approach where learning is continuous.

Conclusion

Nongovernmental organizations (NGOs), governments, and international donor agencies collaborate in partnerships with visions of improving the delivery of services and catalyzing transformative social change. Partner organizations expect benefits such as increased outreach to poor

³¹ Westley, F., Zimmerman B., and Patton M., *Getting to Maybe: How the World Is Changed*, Random House Canada, 2007

³² Zimmerman, B., C. Lindberg and P. Plsek. (1998). *Edgework: Insights from Complexity Science for Health Care Leaders*. Texas: VHA Inc.

³³ Patton, M (2008), *Utilization-Focused Evaluation*, 4th Edition, Sage Publications

communities, improved quality of services through more rapid development and dissemination of ‘best practices’, and greater efficiencies through resource-sharing and coordination of activities. The inherent value of collaboration seems to resonate deeply with board members, senior leaders, and staff members of these agencies, especially when faced with the scale of current social crises. In practice, however, performance can fall short of expectations, at times with such negative consequences that some have begun to question ideas of organizational partnership and collaboration altogether.

This chapter explored some of the definitional issues of partnerships in the not-for-profit sector. Our intent has been to organize our understanding of these complex arrangements in order to provide some reflections on how best to evaluate them. We have presented some observations on the typical characteristics of partnerships and have outlined some of the common challenges that we often face while evaluating these complex entities. We have also proposed some possible ways to address these challenges. Both our observations and our methodological suggestions should be viewed as hypotheses for testing and refinement. It is hoped that they will invite reflection, stimulate discussion, and provide a foundation for nuance with the result that ultimately we shall enjoy a richer understanding of partnerships in the field of International Development and of how to better evaluate them.

CHAPTER FIVE

The Evolution of Institutional and Organizational Assessment

Katrina Rojas and Charles Lusthaus

This chapter reviews ideas associated with the assessment of organizational performance that were discussed with Claremont Graduate University and the Rockefeller Foundation. The Rockefeller Foundation has undertaken an initiative to identify and advance methodologies that can help guide the philanthropy and development community in the evaluation of policy influence, capacity development, networks/partnerships, and organizational performance.

Background and Context

Over the past decades, the evaluation community has focused almost all of its attention on evaluating projects and programs. The Organization for Economic Co-operation and Development – Development Assistance Committee (OECD-DAC), the most common reference point for evaluation criteria, terminology, and standards, has given little guidance on assessing organizations. Their materials refer to development interventions as a general term to indicate the subject of the evaluation, and while this may refer to an activity, project, program, strategy, policy, topic, sector,

operational area, or institutional performance, etc.³⁴, the thrust of the guidelines is the project and program. Thus, in development evaluation practice, the organization as a unit of analysis remains a black box.³⁵

In the United States and Canada³⁶, evaluation practice has also focused largely on evaluating projects and programs. Clearly, from a Universalia perspective the notable exception in Canada has been IDRC and CIDA. Both of these organizations have recognized the importance of assessing organizations since the early 1980s. The majority of sessions at any evaluation conference, such as the Canadian Evaluation Society (CES) and American Evaluation Association (AEA), and listserves and other fora, do not focus on the organization as the unit of analysis. For example, at the 2009 Annual Conference for the AEA, only eight out of 621 sessions (1.3 per cent) discussed issues specifically related to organizational assessment, and most of these focused on some dimension of organizational capacity building.

In contrast to this lack of attention to evaluating organizational performance, donors are increasing their investments in organizations such as government ministries, International Financial Institutions (IFIs), other multilateral organizations, NGOs, and research institutions. Many bilateral and multilateral donors are moving away from project funding to more program and institutional funding, and are raising concerns about the tools available for understanding these organizations to which they are entrusting considerable

³⁴ OECD-DAC (2006). DAC Evaluation Quality Standards. P. 3

³⁵ In physics, a black box is a "system whose internal structure is unknown, or need not be considered."

³⁶ Clearly, from a Universalia perspective the notable exception in Canada has been IDRC and CIDA. Both of these organizations have recognized the importance of assessing organizations since the early 1980s.

funds. Donor agencies that invest in multilateral organizations as a way of fulfilling their missions are increasingly trying to gain a deeper understanding of the performance of these organizations, not only in terms of their contributions to development results, but also in terms of the capacities they have in place to support results achievement.

Similarly, although organizational assessments are being considered as a key tool within the broad continuum of public sector performance management, unlike project and program evaluations, they are not yet institutionalized within government policies. Linking the assessment of the performance of an organization to the assessment of the projects and programs that it carries out seems like a logical step for performance improvement – a step that will catapult the importance of organizational assessment as a useful tool for evaluators.

What is the Assessment of Organizational Performance?

This section explores key definitions. It begins with an overview of the concepts emerging from the literature about organizations and the key elements of their performance and then provides a definition of organizational performance assessment.

Early studies about organizations assumed that they existed to serve a purpose, and that the role of management was to support this purpose by strategically gathering and applying resources in an efficient manner. However, experience showed that organizations did not serve one single purpose,

but had multiple goals and sub-goals³⁷, some of which supported the original 'organizing' purpose, while others did not. Furthermore, those studying organizations suggested that organizations are social constructions, and that managers, staff, and people observing organizations construct their own meaning of what an organization is and how it ought to perform.

In practice, it was found that an organization's goals were constantly being displaced by the actors of the organization. They were displaced in a variety of ways: time changed people's perceptions of the goals, leaders sometimes changed the goals, organizational events caused a shift in priorities, and sometimes changes in the environment, law, or political situation inadvertently acted as counter-productive forces and inhibited the achievement of objectives. Given this complexity, how were organizations and their constituents to know if they were moving in the right direction? How were they to measure performance and the factors associated with good performance?

Caplow³⁸ argued, "every organization has work to do in the real world and some way of measuring how well that work is done". His concept of organizational performance was based on common sense and the notion that organizations need a way of concretely identifying their purpose and assessing how well they are doing in relation to it. According to Caplow, each organization did have a sense of what it was doing and ways of assessing success; in other words, it had an institutional definition of its own purpose.

³⁷ Quinn, R.E., and J. Rohrbaugh.(1983). A Spatial Model of Effectiveness Criteria: Towards a Competing Values Approach to Organizational Analysis. *Management Science* (29): 363-77.

³⁸ Caplow, T. (1976). *How to Run Any Organization: A Manual of Practical Sociology*. Hinsdale, IL: The Dryden Press., 90

Since it was clear to most people and managers that organizations that did not make money went out of business, private firms used the common sense concept of profit as a way to judge their performance. Thus, at the simplest level, measuring profit of an organization was a way of assessing how well the organization was doing. Profit is indeed an important and valid aspect of performance in the private sector, and many managers use profitability as a metaphor for an organization's success. In government and non-profit organizations, however, ideas about what constituted success were much less clear. We all knew that schools helped children learn and that many Foundations wanted to reduce poverty, but there was no root concept equivalent to profit that could be used to assess their success.

Creating methodologies to assess profitability as a primary objective in the private sector was congruent with prevailing ideologies shaping management practices at the time. Management theorists in the early part of the century tended to focus on devising scientific or engineering methods of increasing financial gain³⁹. In support of such management objectives, organizational assessment focused on identifying ways to improve the efficiency of workers. By 'engineering' optimal ways for people to behave in specific organizational production systems, managers aimed to produce more goods for less money, thereby increasing profits.

Starting in the 1940s, more abstract and generic conceptions of performance began to emerge in the discourse on organizational performance⁴⁰. Gradually, concepts such as "effectiveness," "efficiency," and "employee morale" gained ground in the management literature and, by the 1960s, were

³⁹ Taylor, F.W. (1947). *Scientific Management*. New York: Harper and Row.

⁴⁰ Likert, R. (1957). *Some Applications of Behavioral Research*. Paris: UNESCO

considered to be major components of successful organizations⁴¹. Managers understood an organization to be performing if it achieved its intended goals (effectiveness) and used relatively few resources in doing so (efficiency)⁴². In this new context, profit became just one of several indicators of performance. The implicit goal shaping most definitions of organizational performance was the ability to survive. Thus, organizational assessment focused on the extent to which an organization was able to meet its goals within reasonable resource parameters and in doing so make a profit.

Gradually, it became clear that organizational assessment needed to go beyond the measurement of these rather simplistic ideas⁴³. A host of factors emerged as important components to be factored into the assessment equation: the present and future utility of the organization's products and services, productivity, systems, quality, customer satisfaction, innovation, and relevance. Organizational assessment was gradually becoming more holistic, attempting to integrate as many aspects of an organization as possible⁴⁴. Furthermore, assessing organizations was more complex than originally thought. Evaluating or assessing the performance of an organization was anything but a mechanical venture. For example, when Peters and Waterman searched for excellent organizations, they found organizations that met their constructed criteria but it is less than certain that they were in fact excellent organizations, performing well.

⁴¹ Campbell, J.P., et al. (1970). *Managerial Behavior, Performance, and Effectiveness*. New York: McGraw-Hill

⁴² At the time, 'morale' was considered to be a component of broader efficiency indicators.

⁴³ Levinson, H. (1972). *Organizational Diagnostics*. Cambridge, MA.: Harvard University Press

⁴⁴ Levinson, 1972; Gaebler & Osborne, 1993; Harrison, 1987; Meyer & Scott, 1992

It is clear from engaging in such assessments that assessing organizations and their performance is a murky business in the profit world and even more so in the not-for-profit world where there are few industry-wide agreed to criteria of organizational success. Few government agencies, not-for-profit organizations, and foundations go out of business and few of these same organizations have success criteria that they identify, monitor, and use for decision making over time.

What we have found is that at best, we can provide ways of understanding and constructing frameworks and methodologies for helping those adventurous souls in the government, not-for-profit, and foundation world explore their performance through frameworks that can be supported by systematic assessments.

Defining Organizational Performance Assessment

Thus, the question remains, what is organizational performance assessment? Immordino⁴⁵ describes it as “a systematic process for examining an organization to create a shared understanding of the current state of the elements that are critical to the successful achievement of its purpose.” In our own practice, we have defined it as “a systematic process for obtaining valid information about the performance of an organization and the factors that affect performance.” It is a type of evaluation in which the tools of organizational assessment are used to judge the level of performance.

⁴⁵ Immordino, Kathleen M. (2010). *Organizational Assessment and Improvement in the Public Sector*. Boca Raton, FL.: CRC Press, 7

An organizational performance assessment (OPA) differs from other types of evaluations (such as policy, program, and project evaluations) because the assessment focuses on the organization as the primary unit of analysis (i.e. the performance of an organization, not the performance of a project, a program, or a policy). It is conducted to help investors or funders make investment decisions and identify possible areas for improvement. Breaking our definition down into its component parts helps identify the key principles of OPA.

Primarily, OPA is a systematic process or series of actions aimed at answering a set of questions about an organization. It provides a structured framework for collecting, analyzing, and evaluating information that is organizational in nature.

Process

In the OPA context, process means a sequence of steps and a planned methodology. The process used for conducting the assessment is in many ways as important as the results obtained. While the process may vary along a number of dimensions, it always involves people working together to better understand the performance and functioning of an organization within its own context. It provides a way to involve members of the organization in seeking the needed information and may involve them in other components of the assessment processes. The success of the process, and of the relationships created in the course of conducting an assessment, can influence the degree to which the organization will take ownership of the OPA and any changes that may ensue. Thus, the process is inextricably linked to the extent to which the assessment is used.

Systematic

In the OPA context, systematic means having a structured way of collecting information in which decisions are made carefully and conscientiously about the scope and depth of information that is available, how it is to be obtained, and how it will be used.

Stakeholders must have confidence that the information provided in an OPA is trustworthy. It should be collected, processed, and analyzed using the same standards as any other applied social science activity, and should provide the best evidence available to answer the questions posed.

Improving Performance

In addition to knowing how well an organization is performing, we want to know what it can do to improve its performance. Here again, there are a wide range of ideas and concepts (hypotheses, if you will) that can be identified through an OPA as ways to improve performance. For example, is having a clear strategy known to all organizational members a quality that supports improved organizational performance? While there may be any number of changes that an organization can make to try to improve performance, one of the purposes of engaging in OPA is to identify the changes that are feasible. The underlying premise of the authors is that the purpose of organizational assessment is to provide a perspective on the organization that might improve the organization and its performance.

In short, an organizational performance assessment should be an ordered, systematic approach to gathering data on the performance of the organization, its environment, and the

various component parts that support the performance and functioning of the organization.

Frameworks and Models

A number of models and frameworks for organizational diagnosis and assessment have been developed and applied. Many of these link the use of assessment to organizational development, traditionally conceived as an “effort planned organization-wide and managed from the top to increase organizational effectiveness and health through planned interventions in the organization’s processes using behavioural science knowledge.”⁴⁶ An OA thus often feeds into a change process that aims to identify areas in which the organization can initiate improvements in order to reach a higher level of performance, however, that may be defined more effective, more efficient, etc. In this section, we provide an overview of frameworks and models that are recognized as good practices or approaches to organizational assessment. These can be broadly classified into three types of models or frameworks:

- Models/ frameworks that identify best practices or standards associated with strong performance or organizational excellence;
- Models/ frameworks that aim to explore relationships among variables or concepts, some of them based on empirical evidence;

⁴⁶ Original reference is to Richard Beckhard (1969), but this is taken from Immordino (2010:10)

- Results frameworks, which provide an implicit definition of organizational performance that focuses on results achievement.

For more information on these frameworks, refer to the Reflect and Learn web site at www.reflectlearn.org. It is a website dedicated to organizational assessment that has been supported by the International Development Resource Centre (IDRC) and the Rockefeller Foundation.

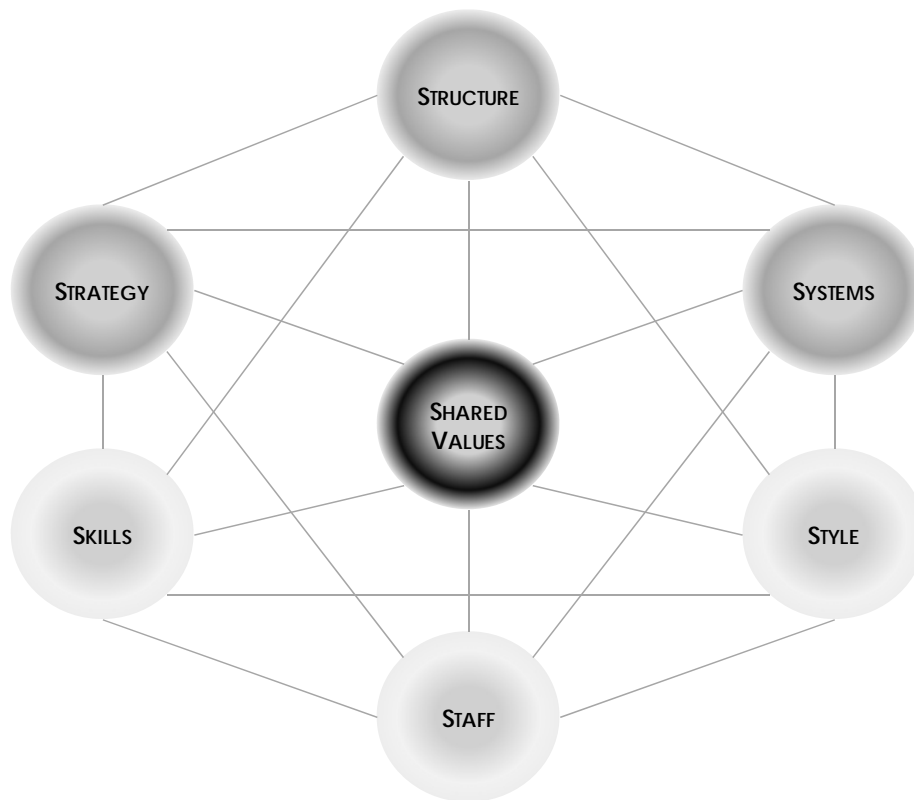
Most models and frameworks that have been used in this field identify a set of organizational practices that are associated with strong organizational performance or organizational excellence. These are often developed into what are referred to as capacity checklists.

Seven-S Framework

The Seven-S Framework (also known as the McKinsey 7-S Framework) was one of the first models of organizational assessment to be popularized almost thirty years ago⁴⁷. It was one of the first to incorporate a holistic or systems perspective in which the interrelationships of key components are seen to determine overall system performance. It was also one of the first popular models to give sustained attention to organizational software, such as human behavioural factors (management style, shared values, staff, skills) as part of a systematic approach to organizational assessment.

⁴⁷ Pascale, Richard Tanner and Anthony G. Athos. (1981). *The Art of Japanese Management*. Elsevier, Business Horizons and Peters, Thomas J., and Robert H. Waterman Jr. (1982). *In search of excellence: Lessons from America's best-run companies*. New York: Warner Books Edition

THE SEVEN-S FRAMEWORK



The original model focused more on activities inside the organization than outside i.e. it did not look at how the environment affected the organization, nor did it address a wide variety of organizational issues such as sustainability, access to financing, power, and control, clients and beneficiaries and many others. Although the model describes the organizational variables and recognizes the importance of the interrelationships among them, it does not explain how each dimension affects the other⁴⁸.

⁴⁸ Burke, W. Warner and Litwin, George H. (1992). A Causal Model of Organizational Performance and Change. *Journal of Management*, Vol. 18, No. 3, 523-545

The Seven S Framework is used widely in the international development community. In adapting the framework, several of these concepts, such as sustainability and relationship to the external environment, have been incorporated.⁴⁹

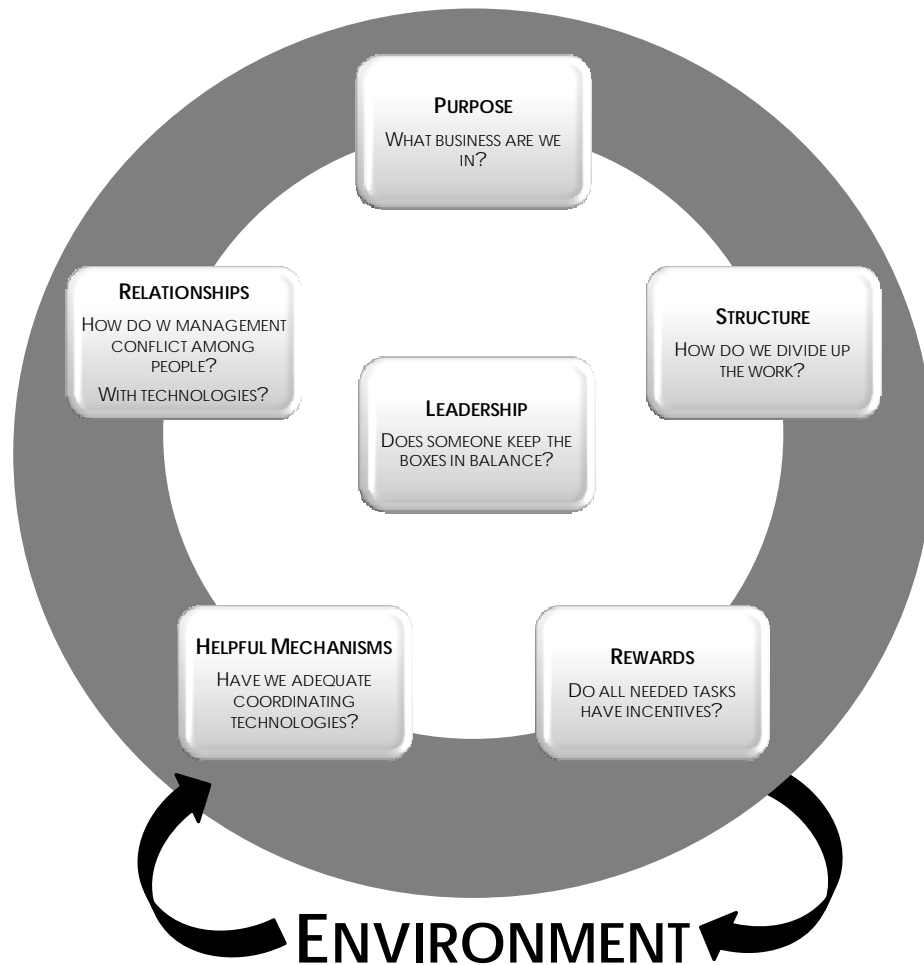
Marvin Weisbord Six-Box Model

The Six-Box Model is a framework that assesses the functioning of organizations, based mainly on the techniques and assumptions of the field of organizational development, and focuses on internal activities of the organization. It is a generic framework and is intended for use across a wide variety of organizations.

The model represents a particular way of looking at organizational structure and design. It gives attention to issues such as planning, incentives and rewards, the role of support functions such as personnel, internal competition among organizational units, standards for remuneration, partnerships, hierarchies and the delegation of authority, organizational control, accountability and performance assessment. The model also follows the basic systems approach to organizational functioning, including the inputs and outputs categories.

⁴⁹ See for example UK Department for International Development (DfID), 2003, Promoting Institutional and Organizational Development: A sourcebook of tools and techniques.

THE SIX-BOX MODEL



Malcolm Baldrige Model

This model, which is the basis for the Malcolm Baldrige National Quality Award (MBNQA) in the United States, provides standards for organizational excellence that can be applied from sector to sector.⁵⁰

⁵⁰ The Malcolm Baldrige National Quality Award was signed into law in the United States in August, 1987. During his term as Secretary of Commerce, Malcolm Baldrige was a proponent of quality management. The award program was designed to stimulate quality improvement processes in private

In general terms, the framework suggests that organizational excellence requires:⁵¹

- Leadership. Examines how senior executives guide the organization and how the organization addresses its responsibilities to the public and practices good citizenship
- Strategic planning. Examines how the organization sets strategic directions and how it determines key action plans
- Customer focus. Examines how the organization determines requirements and expectations of customers and markets; builds relationships with customers; and acquires, satisfies, and retains customers
- Measurement, analysis, and knowledge management. Examines the management, effective use, analysis, and improvement of data and information to support key organization processes and the organization's performance management system
- Workforce focus. Examines how the organization enables its workforce to develop its full potential and how the workforce is aligned with the organization's objectives
- Process management. Examines aspects of how key production/delivery and support processes are designed, managed, and improved

companies as well as the public sector. The Baldrige Award is given annually by the President of the United States to businesses, and to education, health care and non-profit organizations that apply and are judged to be outstanding in seven areas of organizational excellence.

⁵¹ The criteria are adapted in different ways in the literature but generally comprise these elements, which are taken from the National Institutes of Standards and Technology website http://www.nist.gov/public_affairs/factsheet/baldfaqs.htm

- Results. Examines the organization's performance and improvement in its key business areas: customer satisfaction, financial and marketplace performance, human resources, supplier and partner performance, operational performance, and governance and social responsibility. The category also examines how the organization performs relative to competitors.

Models and Frameworks that Explore Relationships between Concepts or Variables

The Causal Model of Organizational Performance and Change, or the Burke & Litwin Model⁵² emerged from the authors' desire to create a guide for both organizational diagnosis and planned management organizational change. It suggests linkages that hypothesize how performance is affected by internal and external factors. It provides a framework to assess organizational and environmental dimensions that are key to successful change and it demonstrates how these dimensions should be linked causally to achieve a change in performance. The model links what could be understood from practice to what is known from research and theory.

The model revolves around 12 organizational dimensions that draw on the variables put forward by other models (Seven S, Weisbord, and others). Although presented here as a simple list, the model is complex and aims to depict

⁵² Burke, W. Warner and Litwin, George H. (1992). A Causal Model of Organizational Performance and Change. *Journal of Management*, Vol. 18, No. 3, 523-545

relationships between variables and explicitly account for variables at different levels of an organizational system, from group and local work unit ideas to individual level:

- External environment
- Mission and strategy
- Leadership
- Organizational culture
- Structure
- Management practices
- Systems
- Work unit climate
- Task and individual skills
- Individual needs and values
- Motivation
- Individual and organizational performance

Universalial/ IDRC Organizational Assessment Framework

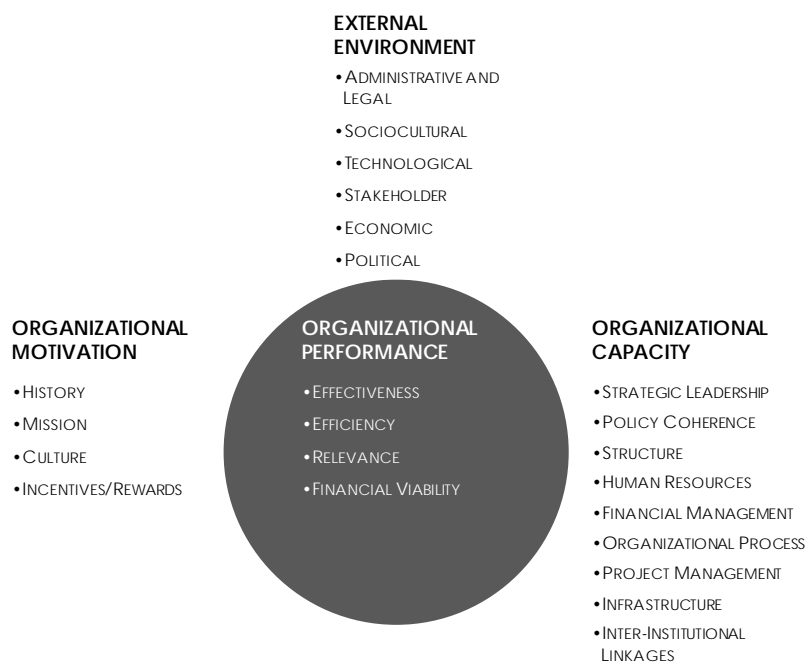
As noted above, Universalial sees performance as a multidimensional idea, defined in our framework as the balance between effectiveness, relevance, efficiency, and financial viability. Each of these concepts is defined and draws on definitions used in the literature.

Organizations do not exist in a vacuum. They are affected by a number of external and internal factors. In designing the Universalial/IDRC framework, we reviewed the literature to identify the factors that seemed to drive change in organiza-

tions. Three key ideas emerged from this review: (i) organizations change because of their external environment, that is the notion of adaptation that is often referenced with regard to drivers of change; (ii) organizations also change as their internal resources (financial, human, technology, etc.) change; and (iii) organizations change when there are fundamental shifts in values that affect climate, culture, and ways of operating.

This exploration led Universalia/IDRC to posit that performance is a function of an organization's enabling environment, organizational motivation, and capacity. The framework aims to understand the relationship between these factors and the organization's level of performance. Only in doing so will it be possible to try to use organizational assessment to improve the organization.

UNIVERSALIA/ IDRC ORGANIZATIONAL ASSESSMENT FRAMEWORK



The framework includes an assessment of capacities that are also considered in the frameworks and models described above, but explicitly adds the concepts of external environment and internal motivation (or organizational culture) as key factors that drive performance in an organization.

The Universalial/ IDRC OA framework explores the relationships between the factors that affect performance and the actual performance of an organization, but there is still limited empirical evidence to help understand if and how those factors affect performance. For example, we assume that having clear strategic plans and inclusive planning processes will have positive effects on performance. Based on experience, we can say that this is a supportive factor, but we do not have the empirical evidence to back this up. Nor do we have the kind of evidence that would help to define the crucial variables in the performance equation for different types of organizations.

Foundation Performance Assessment Framework

Based on research and discussions with foundation leaders in the United States, the Center for Effective Philanthropy⁵³ developed a conceptual framework for Foundation Performance Assessment. The framework seeks to assess the performance of a foundation along multiple dimensions that provides a way of inferring the social benefit created, relative to the resources invested by the foundation. The dimensions include organizational capacity measures and results or effectiveness measures. The framework aims to allow leaders to be able to understand performance in relation to other foundations and over time.

⁵³ The Center for Effective Philanthropy, (2002). Indicators of Effectiveness: Understanding and Improving Foundation Performance. Boston, MA, p.37.)

Models and Frameworks Focused on Results and Metrics

Organizational Results Frameworks

The frameworks noted above emerged from reflection on the nature of organizations and organizational excellence and performance, and the factors that are believed to relate to performance. However, there is another basis for assessing the performance of organizations that emerges from the changing culture of public administration and its increasing emphasis on accountability for results and outcomes.

Through federal legislation in the United States (Accountability Act, 1992) and Canada (Federal Accountability Act, 1996), central agencies are holding departments accountable through departmental and results frameworks. This approach is characterized by measuring progress towards the results that are sought, having the flexibility to adjust operations to better meet these expectations, and reporting on the outcomes accomplished⁵⁴. This has led to growing emphasis on results frameworks – for example the Results Management and Accountability Framework (RMAF) in Canada – as a basis for measuring, reporting on, and assessing organizational performance. In other words, an organization is assessed based on the outputs, outcomes, and impacts that it achieves.

Internationally, there has been a continued emphasis and increased momentum on managing for results. Policies and frameworks in the international context (DAC, Paris Decla-

⁵⁴ Mayne, John. (2001). Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly. *The Canadian Journal of Program Evaluation*, Vol 16 No. 1, pp, 1-24

ration, etc.) provide explanations of the importance of results-based management (RBM) for ensuring that management practices optimize value for money and the use of human and financial resources. Donor agencies that invest in multilateral organizations as a way of fulfilling their missions are increasingly trying to gain a deeper understanding of the performance of these organizations; both their contributions to development results and the capacities in place to support results achievement. From the perspective of these donors, many multilaterals are improving their results frameworks and data-gathering systems, but these are not yet developed enough across organizations to be used as the basis of a systematic effectiveness assessment. As a result, performance information is sought through other means, such as the approach adopted by the Multilateral Organization Performance Assessment Network (MOPAN), which aims to draw on perceptions and secondary data in order to assess organizational performance in terms of systems, behaviours, and practices (or organizational capacities) that are viewed to be important in achieving development results.

The Balanced Scorecard

The Balanced Scorecard⁵⁵ emerged from concerns about performance measurement approaches in the corporate sector that focused exclusively on financial measures. The unidimensional focus was seen to limit an organization's ability to develop a future strategy that could create future economic value. Thus, a scorecard that looked at how an organization was doing from different perspectives –

⁵⁵ Kaplan, Robert S. and David P. Norton. (1996). *The Balanced Scorecard, Translating Strategy into Action*. President and Fellows of Harvard College

initially conceived as financial, customer, internal business process, and learning and growth perspectives – became the pillar of an approach to managing and measuring organizational performance. The Balanced Scorecard provided a framework that could be used for strategy/planning as well as for monitoring and assessment.

The Balanced Scorecard has been one of the driving forces in emphasizing the importance of measuring performance. Indeed, we have seen growing efforts to develop numbers that explain performance and that enable comparison or tracking of performance over time or across organizations. Graham Brown⁵⁶ recognized the need for a more evolved scorecard with metrics that more accurately measure complicated dimensions of an organization's performance. In his critique of the first generation of scorecards, he noted the lack of measurement of ethics and the absence of external factors that could have an impact on the organization's success.

Conducting an OPA

This section provides an overview of the steps involved in an assessment of organizational performance, which include: selecting a conceptual framework, deciding who will conduct the OPA, determining the issues and key questions, identifying appropriate measures/indicators, deciding on data collection methods, and constructing an OPA matrix.

⁵⁶ Graham Brown, Mark. (2007). *Beyond the Balanced Scorecard, Improving Business Intelligence with analytics*. New York: Productivity Press

It is important to begin with a coherent conceptual framework for conducting the OPA and analyzing the OPA data. A number of models or frameworks can be used to conduct an organizational performance assessment. Any of the models described can be used. It is important for the evaluator to consider which framework or combination of frameworks will be most appropriate for the organization that will be assessed.

One of the first considerations in designing an organizational performance assessment is to decide who will conduct the assessment – the organization itself (a self-assessment), an external consultant, or a combination of the two. Each approach has potential advantages and limitations. This decision will have implications for how the process is conducted and who is involved in taking subsequent decisions about the exact methods and measures to use in the process.

Self-assessment is an approach to organizational performance assessment in which the organization itself has considerable control over the assessment. When an organization engages in a self-assessment, it manages the assessment either on its own or with the support of an external facilitator or coach who helps to guide the assessment process. Donors often have input into a self-assessment process, but the assessment questions, data collection, analysis, and report are done by the organization or under the management of the organization.

SELF-ASSESSMENT PROCESSES: STRENGTHS AND LIMITATIONS

STRENGTHS / BENEFITS:	LIMITATIONS/ CONSIDERATIONS
<ul style="list-style-type: none"> ▪ ENCOURAGES OWNERSHIP AND ENGAGEMENT IN A LEARNING PROCESS ▪ PEOPLE IN THE ORGANIZATION HAVE EASY ACCESS TO DATA ▪ ENHANCES SENSE OF DIGNITY AND SELF-RESPECT ▪ INCREASES PERCEPTION OF THE FAIRNESS OF THE PROCESS ▪ INCREASES ACCEPTANCE OF FEEDBACK BECAUSE IT PROMOTES SELF-REFLECTION ▪ INCREASES COMMITMENT TO RECOMMENDATIONS ▪ REDUCES BACKGROUND RESEARCH ON THE ORGANIZATION BY PROVIDING CURRENT INFORMATION ABOUT THE ORGANIZATION AND EXISTING DATA ▪ COST BENEFITS: DOES NOT REQUIRE TIME-CONSUMING PROCUREMENT NEGOTIATIONS, AND MAY NOT NEED TO DRAW ON A SEPARATE BUDGET LINE 	<ul style="list-style-type: none"> ▪ THE INDEPENDENCE OF THE ASSESSMENT MAY BE QUESTIONED BY EXTERNAL STAKEHOLDERS CONCERNED ABOUT THE VALIDITY OF FINDINGS ▪ THE CONCERN THAT HARD ISSUES MAY NOT BE TACKLED ▪ THE PROCESS REQUIRES A GREAT DEAL OF MANAGERIAL TIME ▪ SENSITIVITIES CAN BE STRONG BECAUSE THE PLAYERS ARE INVOLVED WITH THE CONTENT OF THE ASSESSMENT AND HAVE SOME STAKE IN THE ORGANIZATION. A CLEAR DEFINITION OF THE ROLES AND OF THE PROCESS AT THE OUTSET CAN HELP EASE TENSIONS. ▪ THE NOTION OF SELF-ASSESSMENT IS NOT NECESSARILY ACCEPTED IN ALL CULTURES, AND GROUP DISCUSSIONS OF THE ISSUES MAY BE UNCOMFORTABLE FOR PEOPLE IN SOME CULTURES OR ORGANIZATIONS.

The most common situation occurs when donors hire external consultants to conduct independent OPAs. In an external assessment, the donor is responsible for the overall management of the assessment, and both the donor and the organization being reviewed define the basic issues and questions to be explored, but it is the external assessor or evaluator that is responsible for making judgements about the organization’s performance and for ensuring the independence of the final product. While stakeholder input is sought throughout the OPA process, the external consultant determines the methodology and manages the data collection, analysis, and reporting.

EXTERNAL ASSESSMENT PROCESSES: STRENGTHS AND LIMITATIONS

STRENGTHS / BENEFITS:	LIMITATIONS/ CONSIDERATIONS
<ul style="list-style-type: none"> ▪ EXTERNAL ASSESSMENTS ARE OFTEN VIEWED AS MORE INDEPENDENT AND OBJECTIVE ▪ EXTERNAL ASSESSMENTS IMPROVE THE RANGE OF ISSUES ADDRESSED AND THE RELIABILITY OF FINDINGS ▪ THE DONOR CAN SPECIFY REQUIREMENTS FOR CONSULTANT EXPERTISE ▪ EXTERNAL CONSULTANTS: <ul style="list-style-type: none"> ✓ CAN FOCUS EXCLUSIVELY ON THE OPA (AS THEY ARE NOT DISTRACTED BY OTHER ORGANIZATIONAL WORK) ✓ CAN HELP TO SAVE TIME AND HANDLE VERY SENSITIVE ISSUES ✓ ARE MORE LIKELY TO BE AVAILABLE FOR INTENSIVE WORK WITHIN REQUIRED TIMEFRAMES ✓ MAY BRING FRESH PERSPECTIVES AND STATE-OF-THE-ART KNOWLEDGE ▪ EXTERNAL CONSULTING TEAMS ENRICHED WITH LOCAL CONSULTANTS CAN OBTAIN A MUCH BETTER SENSE OF THE ISSUES IN THE SUB-REGIONS, ADAPT DATA COLLECTION TOOLS TO THE CONTEXT, AND LOWER THE OVERALL COSTS OF THE STUDY. 	<ul style="list-style-type: none"> ▪ FEWER OPPORTUNITIES FOR THE ORGANIZATION TO DEVELOP LEADERSHIP/OWNERSHIP FOR THE OPA PROCESS AND RESULTS ▪ REQUIRES MORE TIME FOR CONTRACT NEGOTIATION, ORIENTATION AND SUPERVISION ▪ TIME SPENT ON SITE MAY BE LIMITED BY COSTS ▪ EXTERNAL CONSULTANTS: <ul style="list-style-type: none"> ✓ MAY NOT KNOW THE ORGANIZATION, ITS POLICIES AND PROCEDURES, OR THE AVAILABLE DATA ✓ MAY BE PERCEIVED AS ADVERSARIES, AROUSING UNNECESSARY ANXIETY ✓ MAY NOT BE AWARE OF CONSTRAINTS ON FEASIBILITY OF RECOMMENDATIONS, AND ARE NOT CONTRACTED TO FOLLOW UP ON THEM ▪ EXTERNAL CONSULTANTS REQUIRE THAT PARTNERS INVEST TIME IN SUPPORTING AND ASSISTING THEM DURING THE PROCESS

External consultants should strive to design a participatory approach to the OPA from the outset. There are many benefits to a participatory organizational assessment process. For the organization, the participation of stakeholders throughout the OPA gives them greater opportunities to develop a sense of ownership and commitment to the assessment and its results, and provides greater opportunities for organizational learning, improvement, and even for generating support from other donors. For a donor agency, a participatory OPA gives the donor information that can help it in due diligence and decisions about investing in the partner organization.

Participatory processes can be more time consuming and therefore more costly to implement, so this needs to be factored into the budget for the assignment.

There are also a wide variety of mixed models that integrate different aspects of the assessment approaches described above. For example, an organization may carry out a self-assessment with assistance from an independent evaluator who can provide a suitable methodological framework, coach an internal team to develop appropriate tools, participate in and give advice on data collection and analysis, and review and comment on overall findings and/or reports.

MIXED MODELS FOR ASSESSMENT: STRENGTHS AND LIMITATIONS

STRENGTHS / BENEFITS:	LIMITATIONS/ CONSIDERATIONS
<ul style="list-style-type: none">▪ CAN COMBINE BENEFITS OF SELF-ASSESSMENT (ESPECIALLY ORGANIZATIONAL BUY-IN, ACTIVE PARTICIPATION IN THE OPA PROCESS, AND ACCEPTANCE AND OWNERSHIP OF OPA FINDINGS) WITH THE BENEFITS OF AN EXTERNAL OPA (I.E. INDEPENDENT AND OBJECTIVE, ABILITY TO ADDRESS SENSITIVE ISSUES NOT MENTIONED TO COLLEAGUES OR SUPERVISORS, EXPERIENCE IN DESIGN AND FACILITATION OF OPA PROCESSES).▪ CAN BROADEN THE ISSUES AND PERSPECTIVES COVERED IN THE ASSESSMENT AND THUS INCREASE THE POSSIBILITY THAT FINDINGS WILL BE USED.	<ul style="list-style-type: none">▪ NOT AS INDEPENDENT AND OBJECTIVE AS AN EXTERNAL ASSESSMENT▪ CAN BE MORE TIME CONSUMING THAN EXTERNAL OPA OR SELF-ASSESSMENT AS MORE PLAYERS ARE INVOLVED AND NEED TO BE BROUGHT UP TO SPEED▪ NEEDS TO BE MANAGED CAREFULLY TO AVOID BECOMING A TOKEN PARTICIPATORY APPROACH (E.G., IF INTENT IS TO INCLUDE SELF-ASSESSMENT COMPONENTS, THESE NEED TO BE CAREFULLY PLANNED AND MANAGED).

Determining the Issues and Key Questions

The performance issues or criteria to be explored by the OPA provide the basis for the design of an appropriate methodology. What are the main performance issues in the

organization? The definition of these issues will be based on the conceptual approach that is chosen for the OPA.

EXAMPLES OF PERFORMANCE ISSUES AND KEY QUESTIONS

PERFORMANCE ISSUE	EXAMPLES OF KEY QUESTIONS TO ASK IN AN OPA
<i>EFFECTIVENESS</i>	HOW EFFECTIVE IS THE ORGANIZATION IN WORKING TOWARDS ITS MISSION?
<i>RELEVANCE</i>	HOW RELEVANT IS THE ORGANIZATION TO ITS STAKEHOLDERS?
<i>EFFICIENCY</i>	HOW EFFICIENT IS THE ORGANIZATION IN THE USE OF ITS HUMAN, FINANCIAL, AND PHYSICAL RESOURCES?
<i>FINANCIAL VIABILITY</i>	IS THE ORGANIZATION FINANCIALLY VIABLE?

As in other types of evaluation, there is a need to further break down the key questions into sub-questions that will help to understand these issues. In the Universalia/IDRC framework, for example, there is a compendium of potential questions that relate to each of the elements that comprise performance (effectiveness, relevance, efficiency, financial viability) and the factors that affect the organization's performance (capacity, motivation, and the environment). In preparing an assessment framework, one can refer to this checklist of sub-questions that may be relevant to include but will need to be tailored to the organization assessed.

Once a satisfactory set of questions has been developed to explore the key issues, these should be prioritized according to:

- Resource levels. The time required and resources available to answer the question

- Purpose of the OPA. If the organization has determined that the primary purpose is accountability, for example, then the prioritization might favour questions that will help to respond to this aim.
- Stakeholder interest. Since some questions may be more important to one set of stakeholders than another, the selection of questions will need to reflect a balance of stakeholder needs.

Identifying Appropriate Measures/ Indicators

Performance measures and performance indicators are essentially synonymous. As defined by Harbour,⁵⁷ a performance indicator is a “comparative performance metric used to answer the question, “How are we doing?” along a specific performance dimension and associated performance goal.”

There is no one set of best practice measures or indicators that can be used for assessing the performance of organizations. In an OPA, the important element to keep in mind is that these must be measures of organizational performance. Simply stated, indicators should provide concrete, simple, and reliable measures to answer assessment questions. Indicators can be quantitative or qualitative.

⁵⁷ Harbour, Jerry L. (2009). *The Basics of Performance Measurement*. New York: Productivity Press.

In general, quantitative indicators are numeric representations (e.g. the number of refereed research articles written as a result of a project). Qualitative indicators are less tangible, not always easy to count, and are often based on perceptions of a situation (e.g. descriptions of the ways people found the research useful).

In all the organizations we worked with, the most difficult part of the diagnosis was identifying indicators. Sometimes this was because of an abundance of suggested indicators and the difficulty of weeding out the ones that really answered the assessment questions. Indicators are not the starting point. To develop a good indicator, you first need a clear picture of what you are trying to measure. Then, you will want to consider the relevant data you need, the possible sources of data (there may be many), and the availability and feasibility of getting it. Finally, you may want to consider some of the more subtle implications of indicators:

- Measuring something can give it importance in an organization and can even change organizational activities, for better or for worse. If, for example, the number of experiments conducted by researchers is used as a performance indicator in an OA of a research institution, some may see this as encouraging quantity over quality of research. Alternatively, if the number of people served in a community is used as an indicator in the OA of a social-service agency, this might lead some to believe that they should try to increase the numbers and reduce the time spent with each person.

- Measuring complex dynamics may require a set of indicators. It can be challenging to develop adequate indicators to measure the complex dynamics in an organization. Simple indicators may not always fit the bill and may need to be combined. Most organizations develop a set of carefully considered indicators and modify them over time as they analyze their results.
- Indicators may be interpreted in different ways by different stakeholders. What seems like a straightforward indicator to the assessment team may sometimes point out an organizational paradox or give conflicting signals. For example, an indicator that measures diversification of funding as a measure of financial viability could be seen as both positive and negative. On the positive side, diversification is an indication that the organization is not overly reliant on one donor. However, others may think of this measurement in a different way, as having multiple donors, each with its own priorities, expectations, systems, and evaluation and reporting requirements, can lead to fragmentation and increased costs in managing multiple donor requirements.

In developing indicators for an OPA, it is worth considering some emerging trends with regard to good organizational measures of performance:

- Moving to composite measures of performance. The Balanced Scorecard approach is generating growing organizational experience in the definition of metrics that can be used to measure organizational performance. These metrics, however, often fall short of providing what organizations need to know. For example, most measures on scorecards are measures

of the past (e.g. financial measures), measures that may lack integrity or are too easy to manipulate, or measures that do not provide an accurate reflection of what is really going on (e.g. customer satisfaction or human resource metrics that are still rudimentary). The problem with single measures is that they do not provide enough information to either tell an organization how it is really performing or to diagnose the causes of decline (or improvement) in performance. Thus, Graham Brown⁵⁸ proposes greater use of analytics or composite measures, which focus on a particular aspect of performance and are made up of a series of sub-metrics. The challenge in designing an OPA is to identify what composite measures are important and determine their sub-measures.

- Performance measures across a sector. The research and frameworks of the Center for Effective Philanthropy provide an example of efforts to establish performance measures across a sector. Once such measures are agreed upon, it is important for the sector to come together and to use and continuously confirm that these really are good metrics for organizations in their sector.

Deciding on Data Collection Methods

OPAs follow in the tradition of case study or multiple case study methodologies. A case study is a qualitative form of assessment that draws on both qualitative and quantitative data and relies on multiple sources of information. Organ-

⁵⁸ Graham Brown, Mark. (2007). *Beyond the Balanced Scorecard, Improving Business Intelligence with analytics*. New York: Productivity Press.

izational performance assessments are expected to reflect on multiple sources of data to gain insight into the organization.

We have used the following methods in carrying out organizational assessments:

- Document review. For most OPA questions, some type of documentation is available. This includes reports, file data, memoranda, organigrams, staff lists, policy handbooks, meeting minutes, Board handbook (usually a collection of board policies, roles and responsibilities, etc.), previous studies, audits, assessments, reviews, and so forth.
- Stakeholder interviews. Stakeholders are prime sources of data for OPAs. The evaluators should propose to interview a range of respondents (both female and male) who have knowledge about the organization. These individuals are critical to enhancing the validity of the conclusions that will be drawn.
- Surveys. Surveys are often used to gather data from a large number of people (including internal and external stakeholders). In defining the methodology, it is important to clarify the nature of the survey (objective, target audience), the expected response rate, and any potential limitations to using a survey.
- Site visits/observation. Typically, an OPA requires site visits so that the assessment team can directly observe facilities, physical artifacts, and interactions among staff. Observational evidence can be very helpful in understanding data collected from other sources.

Finally, the components identified above (issues and questions, measures/indicators, and data collection methods) should be articulated in an OPA matrix.

OUTLINE FOR AN ORGANIZATIONAL PERFORMANCE ASSESSMENT MATRIX

ORGANIZATIONAL PERFORMANCE ASSESSMENT MATRIX					
ISSUE	MAJOR QUESTIONS	SUB-QUESTIONS	INDICATORS	DATA SOURCES	DATA COLLECTION METHODS
<i>ORGANIZATIONAL CAPACITY</i>					
<i>ORGANIZATIONAL MOTIVATION</i>					
<i>EXTERNAL ENVIRONMENT</i>					
<i>PERFORMANCE</i>					

Making OA Useful

Using OPAs to improve the performance of an organization is not only beneficial to the organization, it is critical to the future of organizational assessment as a recognized tool for the evaluator’s toolbox. Furthermore, when the results of organizational assessments are under-used, resources are wasted and it throws into question why the assessment took place to begin with.

In this section, we look at what OPAs are used for, and the factors that support the use of OPA results. Ideally, organizations use the results of assessments to help make strategic decisions for the future. They also often need to use assessments and evaluations to justify past behaviour for accountability purposes.

Almost 20 years ago, Universalis explored the utilization of organizational assessments among 69 Canadian NGOs in CIDA’s NGO Division. Many of the results have been replicated in various other settings. The study found that

organizational assessments had two major consumers: the NGO and the funder. It is interesting to note that while the NGO respondents almost always saw the funder as the primary consumer of assessment results, the funders often identified the NGO as the primary consumer. Over 90 percent of the NGO respondents indicated that they used the results of the organizational assessment or were planning to use them.

Planning to use the results of an OPA – Like other evaluation activities, we have learned over the years that it is not enough to complete an OPA report, submit it to the organization and/or client, and assume that utilization will somehow take care of itself. Instead, our lesson is that if we want OPAs to be used, utilization must be planned for from the outset of an assessment and considered throughout the implementation and after reports are submitted and disseminated and debriefed.

At the outset of the assessment, this means deciding on approaches and processes that will ensure the confidence of major stakeholders and support their ownership of OPA results, and also ensuring that sufficient resources are allocated to implement the recommendations. We have found that utilization of OPA results is enhanced when:

- The purpose of the assessment is clear; stakeholders understand the purpose and benefits and agree that the OPA is a valuable exercise.
- The main focus of the assessment is learning rather than accountability.
- Internal leadership is identified for the assessment – leaders or champions within the organization who have the vision, interest, and influence to guide other

people in the organization can be a key to the use of OPA results.

- Stakeholders are involved the OPA process – through involvement in the negotiation and planning stages of the assessment, the commitment of the organization can be developed or enhanced. Participation in the assessment tends to increase the likelihood that the findings will be used.
- Stakeholders see the assessment as relevant, credible, transparent, of high quality, and the findings have validity.
- The assessment team is able to communicate the intent of the assessment, their approach, and the assessment results to senior staff and board members.
- The report is timely – OPA reports that are submitted six months later than expected may not be useful. It is important to consider the planning cycle of the organization.
- There is a process in place and resources allocated for following up on and implementing OPA recommendations.
- Recommendations are realistic and feasible – the financial climate can significantly influence the utilization of OPA results. This means that the financial situation of the organization and the financial implications of the recommendations should be taken into account.

Challenges for Organizational Performance Assessment

As OPA is an emerging area within the evaluator's tool kit, there is no literature that reflects on the challenges. This is in itself a challenge. In this section, we identify challenges that need to be addressed. Some are conceptual, some are methodological, and all are interrelated.

Challenge 1: Importance of organizations and of evaluating organizations

It seems that we need a better debate on the importance and role of organizations in solving social problems. Are organizations a critical part of solving social problems? Should ministries, departments, and government agencies be evaluated as social instruments used to solve social problems? Is the same true for foundations? What about the World Bank? What role do NGOs and other voluntary organizations play? What are the fora for this debate? As noted in the introductory sections of this chapter, there has been limited emphasis on organizations in the field of evaluation. Why is that? Why is it that there have been limited new approaches or contributions to understanding organizations and organizational performance in the last few years?

Challenge 2: Lack of methodologies, measures, and tools to assess complex organizations and a lack of industry standards to make judgments

Over the past ten years, one of the most difficult aspects of assessing organizational performance has been the lack of

progress in developing trustworthy ways of understanding the complexity of organizations. Clearly, there is a wide assortment of methodologies that can be brought to bear on this. Some answer the OPA questions using tools from appreciative inquiry. Others use tools associated with complexity theory. New approaches to results-based management and outcome mapping have provided some insight into methodologies, tools, and measures for assessing organizations. There has been relatively little work done in exploring these approaches and identifying what works and under what circumstances.

For example, Universalis is presently engaged with a network of 16 bilateral donors who are interested in evaluating the multilateral architecture, that is to say all multilateral organizations. The donor governments participating in this network are spending millions of dollars to support organizations whose role it is to solve global economic and social problems. Yet the donors do not feel that they are receiving sufficient information on the performance of organizations such as the World Bank, UNICEF, UNDP, WHO and so forth. For this network, an annual assessment process helps fill the gap in information. The challenge they faced in developing an assessment framework was: what to assess? The group chose the Balanced Scorecard as the framework for their assessment and selected a series of key performance indicators in the areas of operational management, knowledge management, relationship management, and strategic management. The indicators were selected because the group felt they provided indications of the kinds of practices, behaviours, and systems that would enable an organization to perform, and thus ultimately contribute to development results. However, these choices were made in a context in which insufficient work has been done on both the

content and methodology for assessing these complex organizations.

A similar limitation exists with respect to developing industry standards for assessing the effectiveness of many of the organizations we work with. What performance standards should we expect of a foundation, an aid agency, or a research centre? While progress is being made with respect to standards for service organizations that provide health and education and in the philanthropy sector, in many areas the only standards available are those that relate to general management practices e.g. ISO standards, operational audit standards. Thus, we still have limited understanding of what makes a good development NGO, development bank, or humanitarian organization.⁵⁹

We also recognize that there are complex factors in developing industry standards e.g. availability, cost, scope, upkeep, and cataloguing. The challenge today is to increase the types of organizations that have some criteria for determining the quality of their performance.

Challenge 3: Providing concepts that can help organizations respond to demands for greater organizational accountability

Accountability is not new to those involved in evaluation. Throughout most of the 20th century, accountability and performance measurement in the public sector centered on financial accounting, focusing on questions of how much money was spent and on what items, and improved organ-

⁵⁹ For humanitarian organizations, however, the Humanitarian Accountability Partnership has established standards that are used to certify organizations based on their accountability framework, quality of management, and quality of service.

izational performance was defined primarily in terms of efficiency. Today, however, accountability has a broader meaning and includes the results of actions. Performance has come to mean useful products and services, understanding whether constituents are satisfied with the way tasks have been performed as well as efficiency. Current performance assessment efforts are called Managing for Results. This latest management reform attempts to link measures to mission, set performance targets, and regularly report on the achievement of target levels of performance. Organizations attempt to show their constituents what they are getting for their dollars, how efficiently and effectively their tax dollars are spent, and how expenditures benefit beneficiaries.

However, while organizations today talk about their accountability, they rarely report on performance in a way that fully meets expectations for accountability. When one looks at organizational reports, one finds information on some of the organization's programs and projects but not on the organization as a whole. In the not-for-profit world, we lack performance concepts for describing accountability. For example, we have fewer performance concepts to discuss the Rockefeller Foundation (although the Center for Effective Philanthropy framework is an important move in this direction) than we have to discuss General Motors (e.g. in terms of profitability, market share, customer satisfaction, productivity and so forth).

Challenge 4: Adjusting to new organizational types

The global problems facing the world have led many organizations to seek new organizational configurations in order to confront the complexity of the issues they need to address. Among the organizations that develop and imple-

ment assistance programs, some new forms of organizations have emerged, geared to making global development more effective. These include new global movements, coalitions, networks, virtual structures, public-private partnerships (PPP), and so forth.

While the OPA framework has been used to evaluate these new organizational types, the framework has been applied most often to more traditionally structured organizations with defined boundaries and similar life-cycle patterns (i.e. they grow, mature, decline, and eventually pass away). The newer organizational structures are more fluid and dynamic and are continually reinventing themselves. As such, they may have different needs, priorities and sensitivities – which may require a different OPA model.

In addition, some of the processes identified in the OPA framework may need to be assessed in different ways in newer organizational structures. For example, coordination and communication processes are touched on in the OPA framework, but are not central to it. Network organizations, however, are highly dependent on coordination and communication to maintain relationships, and have inherently high transaction costs for these processes. Should efficiency be of paramount importance when trying to solve problems that are inter-organizational in nature? Another issue is that these newer forms of organizations are problem-focused rather than goods and service focused. Does such a focus require a different understanding of performance assessment?

Challenge 5: Heightened concern for organizational social responsibility

While the concept of corporate social responsibility (CSR) began as a way for large private sector organizations to integrate economic, social, and environmental imperatives into their activities, the donor community is gradually incorporating the concept of social responsibility into their operations to influence the processes they seek to change. Should organizations be judged by the level of social responsibility they exhibit?

Today, as citizens and stakeholders are asking organizations to move beyond their typical narrow focus and demonstrate their concerns and responsibilities for society as a whole, there is increasing pressure for organizations to better understand and assess their social responsibility. Where does this idea fit when one explores the performance of an organization? Graham Brown has done some initial exploration of this idea in relation to the Balanced Scorecard. He argues that a new Scorecard Architecture should include the dimension of “leadership/ social responsibility”, based on the fact that ethics and social responsibility are no less important than financial results in terms of their effects on the strong performance, or demise, of an organization. Generally, however, few organizational performance assessment frameworks incorporate such an idea, nor is there any serious exploration of this concept in most of the organizational assessment literature.

In an OPA framework, should social responsibility be added as a performance element, with sub-questions about responsibility for promoting gender equality, environment, human rights, and community development e.g. What is the carbon footprint of the organization? To what extent is it engaged in

practices that support or undermine human rights? How much support does it provide to the community within which it operates? Does it have a coherent way of promoting gender equality? Universalia has worked with clients in trying to incorporate some of these elements into the Universalia/ IDRC framework.

Challenge 6: Having valid data to answer the questions posed by the OPA

Over the years, we have worked extensively with government and various not-for-profit organizations. We have found that they have significantly improved their financial information systems and are now in much better position to explore some issues of efficiency. While some progress has been made, most of our encounters with organizations suggest that their data systems for generating information on their results, their stakeholders, or any other organizational performance issue is still at an early stage of development. For example, even though we have been working in the area of results-based management for almost 15 years, few international organizations have a way of discussing (let alone judging) their outcomes as an organization either quantitatively or qualitatively. Thus, we are always implementing an OPA with sub-optimal data.

This raises an additional challenge in terms of the ethics of making judgments based on incomplete data. We are asked to make judgments about the adequacy of the performance of an organization, but are often dealing with incomplete information.

Conclusion

Building sustainable performing organizations in all spheres is crucial for successful international development, and institutional and organizational assessment can be a key tool towards this end. This chapter has considered the basic concepts in organizational performance assessment, and it has summarized the features of three models and frameworks based on best practices, three models and frameworks that explore relationships between concepts or variables, and two that focus on results and metrics. It has also examined challenges and issues in conducting an organizational performance assessment.

Throughout the examination of these concepts, issues, and approaches, one gets the impression of a dynamic field that is evolving rapidly in an attempt to find useful ways of helping organizations through diagnoses of issues affecting their performance. The contrasting models and approaches are particularly valuable when organizations themselves take ownership of their own learning about how their organization functions and of ways it might improve its performance. New approaches that incorporate improved assessment technology with more standardized protocols for engaging the human drivers of organizational performance is likely a logical next step. Universalia plans to be part of such developments.

SECTION III

*Monitoring and Evaluation
of Specific Thematic/
Sectoral Issues*



CHAPTER SIX

Evaluating Agricultural Systems

Ronald Mackay and Douglas Horton

The origins of agriculture reside in concepts of personal survival, community benefit, and long-term sustainability. Agricultural research has contributed to these goals, but by the 21st century, agricultural research had become an ideological battlefield, pitting against each other those who pursue reductionist, disciplinary science on the one hand and those who advocate a more holistic, multidisciplinary approach on the other. The prize is funding. Given that the two research paradigms have the potential for mutual reinforcement, this adversarial state of affairs is unnecessary. The confrontation has profound implications not only for agricultural research and development (R&D) itself, but also for its evaluation.

This chapter expands on these themes in support of its underlying theme of the state and practice of evaluation in the field of agricultural R&D. We briefly summarise the origins of agriculture while also sketching out the two major positions on the ideological battlefield. We also offer a parallel between this rivalry in research and that within evaluation practice and then step back and review evaluation as an activity in service of the information needs of stakeholders rather than driven by the preferred methodologies of any one discipline. This chapter also looks at some typical current practices in the evaluation of agricultural

R&D drawing on the authors' and others' experience. Later in the chapter, we discuss how the current fractious state of evaluation practice in agricultural R&D cannot be laid solely at the feet of the warring factions. How key stakeholder groups in any specific evaluation undertaking who choose to use, eschew use or even misuse the practice of evaluation and its findings must take some of the blame. We offer instances of use, non-use, and abuse of evaluation with illustrative examples from our own practical experiences and those of others. This chapter concludes with four suggestions on how to restore a more productive relationship within the agricultural R&D community, including its evaluators.

We have provided a broad context that we believe will help the reader's understanding; however, both the context and the account of evaluation practice in agricultural R&D are necessarily abridged and therefore much over-simplified. We have written largely from our own personal experience and have tried to make our points clearly and concisely for those unfamiliar with this particular field.

The Origins of Agriculture and of Agricultural Development

From its earliest days, agriculture has been about people. It represents how they have responded to their changing environment in ways that allow them to survive, organize, evolve socially, and prosper. The agricultural enterprise began as a socially motivated activity directed towards the survival and betterment of the entire group and the whole community.

Around ten thousand years ago, humans in many parts of the world at more or less the same time, began to settle, grow crops and domesticate animals for their social good. Their move to more permanent settlements from nomadic hunting and gathering was made possible by climate change and their evolved intelligence. Hand-in-hand with agricultural production came processing technologies such as the pestle and mortar without which the grains, pulses, root crops etc. that they cultivated would have been inedible. Production and processing together gave humans a competitive advantage over other animals whose level of intelligence allowed them to access only easy-to-find-and-digest foods. Moreover, emerging agricultural societies could create food surpluses. Surpluses allowed groups to become larger and their members more specialised. Agriculture made possible the evolution of modern society.

Over millennia, farmers purposefully selected seeds and bred animals that served them well by providing better and more nutritious returns for the effort they invested. Most of the commodities we enjoy today have their origins in those farmers' selection of the most suitable grains, pulses, and roots and their thoughtful breeding of animals. Their work represents the informal, trial-and-error research and development undertaken for thousands of years within the natural environment.

Food production has become an increasingly complex process. It has evolved to include planning, research, management, environmental protection, operations, storage, processing, distribution, regulation as well as food security, the reduction of hunger, enhanced nutrition, fair access to both food and the necessary agricultural land and the protection of that land. The agricultural enterprise has become a highly complex interrelated system.

Modern Agricultural R&D: An Ideological Battlefield

By the 20th century, agricultural R&D had moved towards a more decontextualized approach impelled by the dominating reductionist approach to science. Scientific reductionism tries to understand the nature of complex systems by examining in detail the relationships among a limited number of their component parts. Much of agricultural R&D moved from farmers' fields and the natural environment into the laboratory and the experimental plot. In these new and controlled surroundings, comparisons between, for example, the production yields of new and improved varieties of seed could be reduced to the manipulation of just one or two factors such as sunlight, moisture, and nutrients. All other variables would be controlled. If control were overly difficult or beyond the framework of the unit of analysis, other variables might simply be ignored.

Early that century, the biologist and geneticist Ronald Aylmer Fisher became the statistician for the Rothamsted Experimental Station in the UK. He developed an experimental approach known as randomised control trials (RCTs) as a way of controlling for biased selection. His *Statistical Methods for Research Workers* was a classic until the end of that century. Randomised control trials are still considered as the gold standard for research by reductionist experimental science. Reductionism was shifting the manner in which agriculture was understood and practised away from the notion of the social good in its broader environmental context towards more direct and short term interests designed to create surpluses of high value commodity crops. This approach would pave the way to increase food production dramatically and thereby enhance food security. Special

interest groups sought surpluses also for the purposes of speculation and self-enrichment, be they governments, private or public bodies or individuals.

However, the gains tended to be neither equitable nor entirely beneficial for the environment. The rural poor benefited less than the more prosperous; marginal land less than areas of high agricultural potential. In many cases, gains were made at the expense of soil and water degradation, reduced biodiversity and long-term sustainability. A minority maintained the pressure to take a more holistic view of agricultural R&D in a variety of approaches including farming systems approach, the ecosystem approach, and the landscape approach driven by agriculture, conservation, and community empowerment.⁶⁰

Integrated natural resource management (INRM) is an approach that integrates research on different types of natural resources into stakeholder-driven processes of adaptive management and innovation to improve livelihoods, agro-ecosystem resilience, agricultural productivity and environmental services at community, ecoregional and global scales of intervention and impact.⁶¹

Recently, Vanloqueren and Baret have characterized agricultural R&D as a battlefield on which holistic science is jostling with a dominant reductionist science. They compare two current technological trajectories to illustrate this battle: genetic engineering and agroecology. Genetic engineering is

⁶⁰ Campbell, B.M.; Hagmann, J.; Stroud, A.; Thomas, R. And Wollenberg, E. (2006) Navigating amidst complexity: Guide to implementing effective research and development to improve livelihoods and the environment. Bogor, Indonesia: Center for International Forestry Research.

⁶¹ Thomas 2002 quoted in (Campbell, B.M. et al. (2006). Navigating amidst complexity: Guide to implementing effective research and development to improve livelihoods and the environment. Bogor, Indonesia: Center for International Forestry Research, 2006.)

“the deliberate modification of the characteristics of an organism by the manipulation of its genetic material” and corresponds to the reductionist scientific paradigm. Agroecology, on the other hand, is “the application of ecological science to the study, design and management of sustainable agroecosystems and corresponds to the holistic scientific paradigm.”⁶²

Within an innovation systems approach, they examine how various factors determine that the lion’s share of research funds is captured by genetic engineering. The prevailing factors active in both the public and the private sectors include, among others: competitiveness; the power of lobbies; the role of the media; the ‘publish or perish’ syndrome; and specialism versus interdisciplinarity. The authors also acknowledge that the ease or difficulty with which agricultural innovations can be evaluated is a factor – the easier to evaluate being preferred over the difficult to evaluate. They conclude that the success of genetic engineering over agroecology has nothing to do with the intrinsic superiority of the former to resolve the challenges facing the planet. Indeed there is growing, well-founded international concern that reductionist agricultural science is not sustainable. The continued success of reductionism, they conclude, is an unplanned, even undesirable result of the interaction of many factors. The result is a ‘lock-in’ situation that promotes genetic engineering and hinders the pursuit of agroecology. A further complication, they suggest, is that proponents of these two paradigms present them as competitors when in fact there is room for collaboration.

⁶² Vanloqueren, G. and Baret, P.V. (2009). How agricultural research systems shape a technological regime that develops genetic engineering but locks out agroecological innovations. *Research Policy* 38, 971-983. Elsevier.

The first generation technologies resulting from genomics raised concerns about the risks of increased spread of known allergens, toxins or other harmful compounds, and horizontal gene transfer particularly of antibiotic-resistant genes and unintended effects. An important consequence is that demand has grown for stronger accountability, stricter regulation and publicly funded evaluation systems to determine objectively the benefits of new sciences and technologies.⁶³

The real challenge is how to link biophysical science and social science. Development problems are not merely scientific; they do not reside at a technological level alone but they are affected by human problems at a political level. Things have to be put into a human context. In addition, we have to build our science undertaking around that human context.⁶⁴

Today, the challenge facing agricultural R&D is how to feed the world population in an equitable manner while protecting the environment against irreversible negative changes. Thus, both the agricultural enterprise and its evaluation inevitably face multidisciplinary challenges, with myriad social, environmental, and economic facets. The broader social and environmental elements are ignored at the risk of missing the meaning, indeed the very sustainability, of the whole. The entire world has learned, during the recent financial crisis that economic models with their inevitable shortcomings and inadequacies are by their very nature much-simplified renderings of a complex and interdependent reality. As such, they provide an inadequate foundation for directing our natural and social systems. For example,

⁶³ Source: IAASTD p. 72

⁶⁴ Source: Interview with Dr. Richard Thomas, ICARDA, 9 October, 2007

the notion of the just, self-regulating, and free market that we have heard so much about has been largely exposed as a myth during the recent world financial crisis.

Alan Greenspan: "The benefits of the market are so great that we just have to pay the price."

George Soros: "Yes, Alan, but the fact is that those who pay the price are never those who reap the benefits."⁶⁵

Nevertheless, into the 21st century the two separate perspectives on agricultural R&D, the reductionist and the holistic, compete. How that competition is resolved will have implications for the future funding, execution and evaluation of agricultural R&D.

However, scientists' views can change. The following anecdote records in his own words how one dedicated biological scientist was led to question the adequacy of his science limited to the laboratory and the experimental plot. Dr. Salvatore Ceccarelli is internationally recognised as one of the originators of decentralised plant breeding, a strategy with important consequences for meeting the livelihood needs of small farmers and the protection of both biodiversity and the environment.

⁶⁵ Source: David Hare (2009), *The Power of Yes*. From Eleanor Wachtel interview with David Hare, *Writers and Company*, CBC, 7 February 2010.

REBUKED BY REALITY: BIOLOGICAL SCIENTIST MEETS FARMER⁶⁶

I am a barley breeder. Our research station was in the wheat agro-ecological zone, not in the barley zone. The wheat breeder got good results and asked me why I couldn't get equally good results from my experimental barley seed. I realized that the barley zone was some distance from the research station and had different agro-ecological conditions.

So my wife, also a barley breeder, suggested, "Why not plant in farmers' fields?"

So I rented a farm in a low rainfall, barley-appropriate area. There was a house on a hill close by where a small farmer lived. I went to the house and asked the farmer to please keep grazing animals off my experimental barley plots. We negotiated an appropriate fee for his services.

Now while I would be working in my experimental barley plots, he would walk down from the house with a pot of tea for me. This was sometimes inconvenient especially if I were counting barley plants in my experimental plots. He would come directly up to me and I would lose count. So I said to him, "It's kind of you to bring tea, but please do not interrupt me. Just wait at the field margin until I have finished counting my row."

So one day, I see him coming with his teapot. He walks to the field margin and waits. I continue counting. Then I see that he has put the teapot on the ground and is going through my experimental barley plots looking at the different varieties.

When I finish counting, he says to me, "I have never before in my life seen so many varieties of barley."

So, when we finish our tea, I say to him, "Walk with me and tell me which varieties of barley you prefer."

I see that he is looking very fast, seeking out the taller varieties. He stops, looks closely at each and then squeezes the stems before selecting: "This one, this one, this one."

He is choosing the very varieties that a barley breeder would not choose – the long-stem varieties tend to lodge, especially in areas where the rainfall is higher. So I ask him, "Why do you choose these? The long-stem varieties lodge." And he says, "This is dry land. They do not lodge." So he knows that each variety he has chosen is suitable for the local conditions.

And I ask him, "Why do you squeeze the stems?" And he says, "I will sell the grain but I need the stems for my sheep and the sheep prefer the softer stems. So the longer stems give more feed."

This is the story of how my participatory plant breeding (PPB) work started.

⁶⁶ Source: Based on an interview between one of the authors (Mackay) and Dr. Salvatore Ceccarelli in Mackay, R. 2008. IDRC's Contribution to the Capacity of ICARDA: a Case Study (second draft).

Dr. Ceccarelli's anecdote shows how one barley breeder trained in the paradigm of biological science, came to appreciate how limited a view of agricultural R&D divorced from its social and environmental context can be. Until then, he had barely considered the knowledge, experience, needs, and preferences of the very farmers for whose benefit his breeding research program existed. The vignette illustrates how far agricultural R&D has strayed from the holistic notion of the common good that was agriculture's driving force for thousands of years. On a more positive note, it also reminds us that some within R&D and its evaluation are turning to a more inclusive appreciation of the broader societal and environmental implications of their work. The research for development of agriculture as a system requires thorough analysis and evaluation of both biophysical and socioeconomic sub-systems and the livelihoods derived from them through agricultural activities.

"The whole is more than the sum of its parts".⁶⁷

Without enlightenment, the ultimate goals of agricultural R&D run the risk of becoming obscured. These goals are founded on principles that seek to promote the betterment of society as a whole and as such go well beyond the merely agronomic or economic. They include reducing hunger and malnutrition; ensuring sustainable management and conservation of water, land, and forests; sustaining biodiversity; promoting opportunities for economic development; facilitating institutional innovation; improving policies; and facilitating institutional innovation.

⁶⁷ Source: Aristotle, 4 BC-322 BC

The Emergence of One Dominant View of the Evaluation of Agricultural R&D

The lion's share of agricultural R&D is conducted within a reductionist paradigm with its few and relatively easily measured variables. This approach lends itself to economic impact assessment. Hence, economic approaches to the evaluation of agricultural R&D have not only tended to evolve parallel to and in synchrony with this dominant reductionist approach but similarly overwhelm other approaches⁶⁸.

Alston, Norton and Pardey, in one of the most prominent guides to the theory and methods of economic evaluation,⁶⁹ present the ascendancy and assumed dominance of economic approaches to the evaluation of agricultural R&D when they announce confidently that the book: 'provides a guide to the theory and methods necessary for evaluating agricultural research and for setting priorities for resource allocation. It reviews, synthesizes, and extends such methods as economic surplus analysis, econometric techniques, mathematical programming procedures, and scoring models. It discusses these practices in the context of scientific policy, describes their conceptual foundations, and explains how to undertake such methods to evaluate agricultural research and assist in research priority setting.'⁷⁰

⁶⁸ Vanloqueren, G. and Baret, P.V. (2009). How agricultural research systems shape a technological regime that develops genetic engineering but locks out agroecological innovations. *Research Policy* 38, 971-983. Elsevier.

⁶⁹ Alston, J. M., Norton, G. W., Pardey, P. G. *Science under scarcity: principles and practice for agricultural research evaluation and priority setting*. CAB International

⁷⁰ Source: CAB International
<http://www.cababstractsplus.org/abstracts/Abstract.aspx?AcNo=19981807923>

Other approaches, i.e. those not strictly quantitative and economic, are summarily dismissed. To all intents and purposes, the evaluation of agricultural R&D has become, in some quarters, synonymous with the statistical measurement of the economic impact of isolated research technologies such as high yielding seed and agronomic practices like monoculture, irrigation, and mechanization.

In essence, economic impact assessment has become to program evaluation what reductionism is to agroecology in agricultural research – the dominant paradigm that attracts the greater share of funding, prestige, and attention. Economic methods are ideally suited to the evaluation of individual research technologies with their (relatively) easily quantified and easily measured variables. Many complex contextual variables are often conveniently ignored in economic models.

One of the biggest problems with economics is that economists focus their attention almost entirely on things that can be measured. Moreover, those things that can be measured most easily and accurately tend to get more attention than those more difficult to measure.⁷¹

This disregard for the wide range of approaches to program evaluation other than the economic has implications not only for creating a view of what counts as evaluation but also for influencing what is to be counted as legitimate agricultural research. The evolving circular argument can be phrased as: The evaluation of agricultural R&D is best undertaken using economic methods; agricultural R&D that is most deserving of funding is that which can be evaluated using economic methods. The randomized control trial

⁷¹ Source: Bartlett,
http://findarticles.com/p/articles/mi_qa3827/is_199904/ai_n8845226/pg_1

approach has also been added as a touchstone to assess the fundability of agricultural R&D.

What is Evaluation About?

Before addressing the implications of the dominance of an economic approach to evaluation, we feel it necessary to review evaluation practice from the perspective of the broad community of professional program evaluators as opposed to limiting it to the standpoint of a single discipline.

Rather than accept as dogma that economic evaluations are superior, we suggest what we consider a more empirical and thoughtful approach to the matter. Surely one evaluation approach or methodology cannot be intrinsically superior to another. In any given context, however, one approach is likely to be better suited to answer specific questions that puzzle specific stakeholder groups. One methodology rather than another may be better suited to address particular questions, to collect the kinds of data that are most pertinent and that stakeholders find most credible. One approach rather than another may allow evaluators, together with stakeholders, to draw on data sets to arrive at enlightened answers; to offer options for reporting that specific groups find understandable, illuminating and helpful in practical decision-making.

Writers on methods tend to categorise evaluation into different approaches: One of the authors and a number of collaborators published a comprehensive sourcebook on the evaluation of agricultural R&D in 1993.⁷² It included an

⁷² Horton, D., P. Ballantyne, W. Peterson, B. Uribe, D. Gapsin and K. Sheridan. 1993. *Monitoring and evaluating agricultural research: a sourcebook*. CAB International: Wallingford, U. K.

overview of M&E principles and processes as well as recommendations on how to design and carry out an evaluation. It also included a wide range of methods and examples of their use.

Conceptually useful as these categorizations may be, they have the potential for misleading the casual – and even the more sophisticated – reader into assuming that choosing an approach is the starting point of an evaluation. Our experience and our understanding of the state of the field assert that the fundamental purpose of evaluation is to answer questions. We believe that it is the evaluation questions that are the priority. Questions reflect the concerns of stakeholders. How these questions are answered determines the type of evaluation to be undertaken, the approach and methodologies to be adopted in any given instance. Evaluations must be driven by stakeholders' questions not by evaluators' methodologies. This is the principle of putting stakeholders' information needs first. Patton calls it the 'personal factor'.⁷³

This personal factor holds that evaluations are undertaken to answer the questions of identified stakeholders who need information and who intend to use it to make decisions about policy change, project improvement, funding, or some other substantive decision.

By contrast, the choice of an approach as the first step necessarily limits the questions that can be asked to only those that the methodology is capable of answering. Hence, starting out with an economic outcome or any other approach breaches the personal factor – the principle that gives priority to stakeholder questions. Evaluation needs to

⁷³ Patton, M.Q. (2008). *Utilization-focused evaluation*, 4th edition. Thousand Oaks, CA: Sage.

“choose the methods that are most appropriate for the questions being addressed, rather than because of the familiarity or disciplinary background of the researchers”.⁷⁴

Typical stakeholders in the delivery and evaluation of agricultural R&D projects include the poor and the hungry, national governments, policy makers, research funding agencies, scientists, the tax-paying public, farmers, the public service, community developers, lending banks, conservationists, development agencies, faculties of agriculture, animal nutrition and forestry; national research institutes. All who stand, in one way or another, to suffer if R&D results are poor and to benefit if they are sound are legitimate stakeholders.

It will be relatively easy to respond to some of the important questions that one or other of these groups may pose and want to have answered. It may take the efforts of evaluators trained in only a single discipline. For example, it is relatively straightforward for an economist to answer questions about phenomena that are simple to measure and for which quantitative databases may already exist – e.g. which variety of barley produces higher yields in a particular agro-ecological zone? How many farmers have adopted a new variety of barley? How many hectares of the new variety are being planted? What additional production of barley has resulted from adoption of the new variety? To what extent have farmers’ incomes increased as a result of adopting the new variety?

⁷⁴ Meinzen-Dick, R.; Adato, M.; Haddad, L.; Hazel, P. (2003). Impacts of Agricultural Research on Poverty: Findings of an Integrated Economic and Social Analysis. Washington DC; IFPRI.

STAKEHOLDER GROUPS, EXAMPLES OF THEIR PRIMARY INTERESTS AND POTENTIAL QUESTIONS⁷⁵

STAKEHOLDER GROUP	INTERESTS	POTENTIAL QUESTIONS
<p><i>THE FUNDER/PROMOTER</i> (<i>GOVERNMENT; QUASI-GOVERNMENT DEVELOPMENT AGENCIES; DEVELOPMENT BANKS; TAXPAYERS; CITIZENS WHO DONATE TO NGOs</i>)</p>	<ul style="list-style-type: none"> ▪ SERVICE TO THEIR COUNTRY, SERVICE TO LESS FORTUNATE COUNTRIES ▪ SOCIOECONOMIC IMPROVEMENT ▪ ASSURANCE THAT THEIR TAXES ARE BEING WELL SPENT; CONFIRMATION THAT THEY ARE GETTING VALUE FOR MONEY ▪ ASSURANCE THAT THEIR LOANS/DONATIONS ARE PRODUCING ANTICIPATED RESULTS; VALUE FOR MONEY ▪ FUNDING PROJECTS THAT FURTHER THEIR MISSION; CONFIRMATION THAT THEY ARE GETTING VALUE FOR MONEY ▪ CONFIRMATION THAT R&D IS SUSTAINABLE; THAT IT RESULTS IN NO IRREVERSIBLE DAMAGE TO THE ENVIRONMENT 	<ul style="list-style-type: none"> ▪ WHICH POLICIES SHOULD WE PURSUE? ▪ ARE OUR POLICIES SOUND? ▪ HOW CAN OUR POLICY ON X BE IMPROVED? ▪ IS OUR TAX MONEY BEING WELL SPENT BY OUR GOVERNMENT? BY OUR NATIONAL DEVELOPMENT AGENCIES? BY THE INTERNATIONAL LENDING BANKS OUR GOVERNMENT PROVIDES DONATIONS TO?
<p><i>THIRD PARTY BENEFICIARIES</i> (<i>AID RECIPIENTS; FARMERS ETC. IN DEVELOPING COUNTRIES</i>)</p>	<ul style="list-style-type: none"> ▪ FOOD SECURITY, NUTRITION, HEALTH, DIGNITY, IMPROVED LIVING CONDITIONS, POVERTY REDUCTION 	<ul style="list-style-type: none"> ▪ ARE OUR INTERESTS AND CONCERNS BEING TAKEN INTO CONSIDERATION WHEN RESEARCH PROJECTS ARE PLANNED, FUNDED AND EXECUTED? ▪ DO WE ENJOY THE ANTICIPATED BENEFITS OF THE PROJECT? DO WE LOSE ANY OF OUR EXISTING BENEFITS?
<p><i>EXECUTING AGENCY</i> (<i>GOVERNMENT OR NON-GOVERNMENT AGENCY WHO ACTUALLY MANAGES PROJECTS</i>)</p>	<ul style="list-style-type: none"> ▪ DOES OUR ROLE IN THE PROJECT FURTHER OUR ORGANIZATIONAL STRATEGY, GOALS AND OBJECTIVES? 	<ul style="list-style-type: none"> ▪ ARE THE ASSUMPTIONS AND THE THEORIES UPON WHICH THE PROJECT IS BASED, ADEQUATE? ▪ IS THE NATURE OF THE OUTCOMES AND RESULTS THAT WHICH OUR ORGANIZATION SEEKS?
<p><i>PROFESSIONALS AND PARTNERS</i> (<i>THOSE WHO PARTNER WITH THE EXECUTING AGENCY TO IMPLEMENT THE PROJECTS</i>)</p>	<ul style="list-style-type: none"> ▪ DOES OUR ROLE IN THE PROJECT FURTHER OUR ORGANIZATIONAL STRATEGY, GOALS AND OBJECTIVES? 	<ul style="list-style-type: none"> ▪ ARE THE ASSUMPTIONS AND THE THEORIES UPON WHICH THE PROJECT IS BASED, ADEQUATE? ▪ IS THE NATURE OF THE OUTCOMES AND RESULTS THAT WHICH OUR ORGANIZATION SEEKS?

⁷⁵ Source: based on Lempert, D.H. (2010). Why Governments and Non-Governmental policies and Projects fail despite "Evaluations": An Indicator to Measure whether Evaluation Systems incorporate the Rules of Good Governance. Journal of MultiDisciplinary Evaluation, Volume 6, Number 13. <http://www.jmde.com/>

However, evaluators may find it very difficult to answer other, equally important and legitimate questions. Questions in the difficult category are those that ask about the processes involved in bringing about outcomes and the causal mechanisms between them that produce development results. Complex questions may evade precise answers. Even barely adequate answers to complicated questions may require the collaboration of specialists from several disciplines using multiple methods and drawing on both quantitative and qualitative data. Irrespective of the level of difficulty or interdisciplinary collaboration involved, evaluators cannot discourage or duck difficult questions. Their task is to provide the best answers they can, presented in a clear and useful way. Even if they cannot be answered with precise numerical values then ways must be sought to answer them at least to a level of precision that allows those who need the information to continue to make decisions with the best information available. If evaluation studies are not seen merely as one-off efforts they can be managed as valuable and cumulative data banks that serve as repositories for organizational knowledge and help organizations to grow in effectiveness by learning as they approximate answers to their important questions about how successful R&D is brought about.

It is misguided to assume that the only legitimate questions that can be asked about agricultural R&D projects are those that can be answered by a single discipline or by using quantitative methods or by employing RCTs. An equally grave mistake would be to believe that the only legitimate methods in the evaluation of any R&D project are approaches that employ experimental methods or economic analyses. It is neither moral nor legitimate to limit stakeholders' questions only to those that can be easily answered by RCTs and statistical analyses. Such misguided precepts

could result in catastrophic consequences for the conceptualisation and formulation of agricultural R&D projects. Agencies that fund agricultural R&D could eliminate from their consideration necessary and important R&D work if they were to limit their support only to projects that lend themselves only to RCTs or economic impact assessment and refuse to fund projects whose results were hard to measure or could not be reduced to numbers or required creative interdisciplinary approaches.

What are the Implications of this Conflict?

An adversarial relationship has evolved between two groups of evaluators who conduct evaluations of agricultural R&D. In one camp are those who see economic impact assessment as the necessary approach to the evaluation of agricultural R&D. In the other are those who believe that the approach selected in any given evaluation should be based on the context, stakeholders' questions, and how best these questions can be answered to satisfy what the stakeholder needs to do with the answers. We use the term 'responsive' for the latter perspective.

Despite an apparent unwillingness or inability to meld these stances into a broader and more comprehensive view of evaluation as a field of practice with options that are contextually determined, each group nevertheless works productively to refine useful approaches within their separate ranges of expertise.

Economic methods as drivers of impact assessment

The economic perspective on agricultural R&D is that investments in agricultural research generate high returns and are central to economic growth. Economic growth is the main vehicle for reducing poverty; therefore, growth in the agricultural sector plays a major role in overall economic growth and is a factor in connecting the poor to growth. Hence, investment in agricultural R&D is a key to poverty reduction. The economic lines of reasoning were pulled together in a seminal book, *Transforming Traditional Agriculture*. It presents an elegant case for the central role of agricultural R&D to supply modern factors of production in agricultural growth. Schultz won the Nobel Prize in Economics in 1970. The strictly economic rationale for investing in agricultural R&D played a key role in convincing donors to support the establishment of the Cooperative Group on International agricultural Research (CGIAR), a strategic alliance of members, partners, and international agricultural centers that mobilizes science to benefit the poor.

Recently, Prahbu Pingali, then Director of the Economics Program at CIMMYT and now deputy director of Agricultural Policy and Statistics in the Bill and Melinda Gates Foundation, published a paper entitled “Milestones in impact assessment in the CGIAR.” In this paper he celebrates his admittedly insider’s point of view covering three decades, the “enormous contributions by CG economists to the science of impact assessment.” He highlights the fact that on numerous instances, the CGIAR has been a forerunner to a substantial body of academic research literature on particular themes related to economic impact assessment.

Responsive program evaluation

Program evaluation is really a superordinate term to refer to systematic methods for collecting, analyzing, and presenting information to answer stakeholders' questions about projects, policies, and programs in ways that they find clear, credible, and useful. This use of the term is compatible with how most of the professional evaluation groups view evaluation. Professional expressions can be found at the websites of the Canadian Evaluation Society⁷⁶, the American Evaluation Association⁷⁷, and others.

Multi-methods evaluation

As the CGIAR moved from productivity increases to poverty reduction as its goals, it commissioned a study led by the International Food Policy Research Institute (IFPRI). While some components of this study used conventional econometric analyses, a collection of five case studies used a combination of qualitative and quantitative techniques to address the social as well as economic impacts. Their and others' experience with agricultural R&D indicated that anticipated benefits predicted by a simple casual pathway from agricultural research to increased food production and thereby to poverty reduction do not always materialize. Moreover, some effects from R&D can result in some of the poor being worse off. The IFPRI group notes that economic models that simplify the linkages between agricultural R&D, increased yields, and poverty inevitably limit the questions that researchers as well as stakeholders can ask in evaluation studies. As a result, many questions that ask about impor-

⁷⁶ <http://www.evaluationcanada.ca/>

⁷⁷ <http://www.eval.org/>

tant aspects of how poverty may best be alleviated – or how it may be exacerbated – are ignored because they cannot be accommodated within the models. To understand the complex linkages this group has shown considerable ingenuity and creativity with multi-method approaches that are better capable of addressing difficult questions and acknowledging the complexity of the issues than any lone method might.

They have adopted a livelihoods framework for their studies. One of the advantages of this conceptual framework is that it allows them to take a complex range of interrelated social, economic, agronomic, and institutional factors into consideration. These factors include all of the capabilities, assets and activities that farmers and their families need to provide a means of living as well as threats to the environment; factors which give rise to vulnerability; the contextual policies, institutions and processes; and the strategies pursued by farmers, their families and communities to provide a means for living. Their impact evaluation work suggests that to more completely understand the role of agricultural research in the lives of the poor, still additional aspects of culture, power, and history need to be integrated within the framework. Their work points out the inadequacy of a simple causal pathway by means of which agricultural research disseminates its benefits.

The sustainable livelihoods framework provides a common conceptual approach to examining the ways in which agricultural research and technologies fit (or sometimes do not fit) into the livelihood strategies of households or individuals with different types of assets and other resources, strategies that often involve multiple activities undertaken at different times of the year. Applying this

framework requires interdisciplinary research and a combination of quantitative and qualitative methods.⁷⁸

Recently the approaches pioneered by the group have been used to complement more macro-level economic impact assessments showing a promise for sociologists and economists to work productively together to produce better evaluations than either could alone.

Many of the concerns shown in their work, including those for finding more adequate evaluation frameworks and methods to respond to the essential complexity of agricultural R&D, are repeated by the IAASTD. That body notes that the evaluation of agricultural research and development currently faces both a challenge and an opportunity that require special efforts and investments if it is to answer questions relating to all of the relevant factors including “climate change, land degradation, reduced access to natural resources (including genetic resources), bioenergy demands, transgenics, and trade”.

Social equity issues, including gender, are major concerns in agriculture, as they relate to poverty, hunger, nutrition, health, natural resource management and environment, which are affected by various factors resulting in greater or lesser degrees of equity.⁷⁹

⁷⁸ Adato, M., and R. S. Meinzen-Dick (eds.). 2007. *Agricultural research, livelihoods, and poverty: Studies of economic and social impacts in six countries*. Baltimore, MD: Johns Hopkins University Press and International Food Policy Research Institute; New Delhi: Oxford University Press.

⁷⁹ Beverly D. McIntyre et al. (eds) (2009). *International Assessment of Agricultural Knowledge, Science and Technology for Development International assessment of agricultural knowledge, science and technology for development (IAASTD): global report*.

Questions as drivers of evaluation

Essentially, evaluation is a reality check for those who plan, fund, execute, and hope to benefit from agricultural R&D. The fundamental task of the evaluator is to answer stakeholders' questions in clear and useful ways; to reduce uncertainties by answering the questions that those who stand to gain or lose by development projects are entitled to ask and to have answered so that they can make informed decisions about their assumptions, activities, current results, and future actions.

No constraints on the questions asked can be tolerated. Different groups are entitled to ask any questions about any aspects of projects in which they have a stake. All are legitimate if genuinely generated out of stakeholder needs. The task of the evaluation team is then to plan an evaluation approach and adopt methods that will respond to these questions in ways that serve well those who need the answers.

In an evaluation of a project undertaken to develop the capacity of national agricultural research organizations to plan, monitor and evaluate their research, the authors were asked to address three broad questions, geared to the information needs of the executing agency (ISNAR), the beneficiaries (NARS and their scientists) and the funding agency (the Inter-American Development Bank):

- What were the main contributions of the project to agricultural research management?
- How were the project's contributions brought about?

- What lessons can be learned to improve the design of future capacity development programs?⁸⁰

The rationale for the project and the occasion of the evaluation was acknowledgement that a lack of capacity to manage research and development organizations effectively presents a major obstacle to the potential for agricultural research to contribute to human welfare, food security, and sustainable environmental management. Managers need to monitor activities and evaluate results in order to adjust, redirect, and improve the effectiveness of their organizations' efforts. They also need to learn, from post-hoc evaluations, about the strengths and weaknesses of these efforts. ISNAR launched a three-year project to help develop capacity in national agricultural research organizations in Latin America and the Caribbean. The objective was to strengthen the planning, monitoring, and evaluation (PM&E) capabilities of managers in agricultural research organizations and thereby enhance R&D management.

In order to achieve its goals, the PM&E project carried out activities in three areas:

- The provision and dissemination of contextually-relevant information. Reference books and training materials were prepared for use in training events and workshops and for distribution to managers and libraries throughout Latin America.
- Training and workshops. A regional group of trainers was established, and its members organized and delivered a number of sub-regional training events.

⁸⁰ Horton, D., R. Mackay, A. Andersen, and L. Dupleich. 2000. Evaluating capacity development in planning, monitoring, and evaluation: A case from agricultural research. ISNAR Research Report No. 17. The Hague: International Service for National Agricultural Research.

- Facilitation of organizational change. The project provided direct expert support for organizational change processes in selected organizations that were committed to making significant improvements in their PM&E systems. These pilot cases were in Costa Rica, Cuba, Panama, and Venezuela.

EVALUATION MATRIX: THE FIVE COMPLEMENTARY EVALUATION STUDIES

STUDY	OBJECTIVES	METHODS	SOURCES OF DATA
STUDY 1 <i>THE ISNAR PM&E CAPACITY DEVELOPMENT PROJECT</i>	REVIEW THE PROJECT'S OBJECTIVES, STRATEGIES, ACTIVITIES AND OUTPUTS	SELF-ASSESSMENT	PROJECT RECORDS
STUDY 2 <i>IMPACTS OF INFORMATION</i>	ANALYZE DISSEMINATION, USE AND IMPACT OF PUBLICATIONS	MAIL SURVEY	500 RECIPIENTS OF PROJECT PUBLICATIONS FROM 140 ORGANIZATIONS IN 24 COUNTRIES
STUDY 3 <i>IMPACTS OF TRAINING</i>	ANALYZE IMPACTS OF TRAINING	MAIL SURVEY	150 TRAINING PARTICIPANTS FROM 60 ORGANIZATIONS IN 20 COUNTRIES
STUDY 4 <i>CHANGES IN PM&E IN THE PILOT CASES</i>	ANALYZE CHANGES IN PM&E IN THE PILOT CASES; IDENTIFY CONTRIBUTIONS OF THE PM&E PROJECT; DETERMINE EFFECTS OF THE CHANGES ON ORGANIZATIONAL PERFORMANCE	FACILITATED SELF-ASSESSMENT	COLLABORATORS IN 3 PILOT CASES
STUDY 5 <i>DYNAMICS OF PM&E IN LATIN AMERICA AND THE CARIBBEAN</i>	ANALYZE CHANGES IN PM&E IN THE REGION; IDENTIFY CONTRIBUTIONS OF THE PM&E PROJECT; DETERMINE EFFECTS OF THE CHANGES ON ORGANIZATIONAL PERFORMANCE	CASE STUDIES	INFORMANTS, DOCUMENTS AND OBSERVATIONS IN 9 ORGANIZATIONS IN 8 COUNTRIES

Five evaluation studies were designed together with key stakeholders to provide information that would answer these questions in a credible and useful manner. The evaluation matrix shows the objectives of each study and the

methods employed to collect the necessary data from the appropriate sources. The data sets developed included mail surveys, self-assessments both facilitated by the evaluation team and undertaken entirely by the informants, and case studies that involved personal observation, document analysis and structured interviews.

Utilization-focused evaluation⁸¹

Utilization-focused evaluation (UFE) in agricultural R&D builds on a long tradition of participatory and collaborative monitoring and evaluation. In late 2006, the International Network for Bamboo and Rattan (www.inbar.int) engaged one of the authors (Horton) to evaluate its programs. Headquartered in Beijing, INBAR's mission is to improve the wellbeing of bamboo and rattan producers and users while ensuring the sustainability of the bamboo and rattan resource base. The Dutch Government had initially requested and funded the evaluation as an end-of-grant requirement.

- Step 1. Identify primary intended users. The first task was to ascertain the real purposes and potential users of the evaluation. This process began with a face-to-face meeting with INBAR's Director and a call to a desk officer at the Dutch Ministry of Foreign Affairs, which revealed that the intent of both parties was for the evaluation to contribute to strengthening INBAR's programs and management. During an initial visit to INBAR's headquarters, additional stakeholders were identified, including INBAR board members and local partners.

⁸¹ Source: Patton, M.Q. and Horton, D. (2009). Utilization-focused evaluation for agricultural innovation. ILAC Brief No. 22. The ILAC Initiative, Bioversity, Rome.

- Step 2. Gain commitment to UFE and focus the evaluation. From the outset, it was clear that key stakeholders were committed to using the evaluation to improve INBAR's work. Therefore, the main task was to identify key issues for INBAR's organizational development. Three methods were used: (1) a day-long participatory staff workshop to review INBAR's recent work and identify main strengths, weaknesses and areas for improvement; (2) interviews with managers and staff members; and (3) proposing a framework for the evaluation that covered the broad areas of strategy, management systems, programs and results.
- Step 3. Decide on evaluation methods. After early interactions with the Dutch Ministry of Foreign Affairs on the evaluation TOR, most interactions were with INBAR managers, staff members and partners at field sites. It was jointly decided that INBAR would prepare a consolidated report on its recent activities (following an outline proposed by the evaluator) and organize a self-evaluation workshop at headquarters. The evaluator would participate in this workshop and make field visits in China, Ghana, Ethiopia and India. INBAR regional coordinators proposed schedules for the field visits, which were then negotiated with the evaluator.
- Step 4. Analyze and interpret findings and reach conclusions. At the end of each field visit, a debriefing session was held with local INBAR staff members. At the end of the field visits, a half-day debriefing session and discussion was held at INBAR headquarters; this was open to all staff. After this meeting, the evaluator met with individual staff

members who expressed a desire to have a more personal input into the evaluation process. Later on, INBAR managers and staff members were invited to comment on and correct a draft evaluation report.

- Step 5. Disseminate evaluation findings. The evaluator met personally with representatives of three of INBAR's donors to discuss the evaluation's findings, and the final report was made available to INBAR's donors, staff members and the Board of Trustees. A summary of the report was posted on the INBAR website.

The evaluation process helped to bring a number of issues to the surface and explore options for strengthening INBAR's programs. For example, one conclusion of the evaluation was that INBAR should seek to intensify its work in Africa and decentralize responsibilities for project management to the region. There has been a gradual movement in this direction, as new projects have been developed. INBAR has recently opened a regional office for East Africa in Addis Ababa and is putting more emphasis on collaboration with regional and national partners.

Participatory Impact Pathways Analysis (PIPA)

PIPA is a modus operandi that draws from and expands upon existing evaluation approaches and methods.⁸² It engages stakeholders in a structured participatory engagement. It promotes learning and provides a common framework for action research on processes of change in real world contexts. PIPA is being increasingly used as a practi-

⁸² Douthwaite, B. (2008). Participatory Impact Pathways Analysis. ILAC Brief No. 17. ILAC, Bioversity, Rome.

cal approach to the planning, monitoring and evaluation of complex agroecological projects. It can also be employed for ex-ante and ex-post impact assessment.

PIPA involves the participation of all stakeholder groups.⁸³ It begins with a workshop where together, stakeholders define how they believe that their project can achieve the results they seek. Using construct problem trees, visioning exercises and network maps, they define, refine, and clarify the impact pathways that field experience and existing knowledge deem most likely to succeed. These detailed impact pathways are then used to build two separate logic models. The first, the outcomes logic model describes the project's medium term objectives in the form of hypotheses: Which actors need to change their behaviour? What changes must occur? What strategies and activities are most likely to realise these changes? The second, the impact logic model describes how, by helping to achieve the expected medium term outcomes, the project will affect people's livelihoods. Participants then derive outcome targets and milestones, which are regularly revisited and revised as part of project monitoring and evaluation.⁸⁴

⁸³ More information on all aspects of PIPA, including an on-line manual, can be found at <http://impactpathways.pbwiki.com>.

⁸⁴ Douthwaite, B.; Alvarez, S.; Thiele, G and Mackay, R. (2008). Participatory Impact Pathways Analysis: A practical method for project planning and evaluation. ILAC Brief No 17. The ILAC Initiative, Bioversity, Rome.

Observations on Use, Non-Use and Misuse of Evaluation

The numerous groups who have a stake in evaluations have the potential to present intentional or unintentional challenges to sound evaluation practice.

In any given evaluation study, stakeholder groups tend to overlap. For example, a donor agency funding the R&D may (i) prepare the terms of reference for an evaluation and put up the funds for an evaluation; (ii) commission an external individual or team to plan and execute it; (iii) form part of a group of stakeholders who interprets the data collected by the evaluation team and draws recommendations from the data; (iv) take the evaluation results and/or recommendations into account when making future decisions about their programs and funding priorities; and (v) find its funding expanded or cut by those who finance its activities based on the success of its investments.

Another example might be when a program manager (i) commissions a mid-term external evaluator to evaluate his or her program as part of the original contractual conditions laid down by the funding agency. Then the manager and the evaluation team, together or with a group of internal and external stakeholders may (ii) decide on the evaluation questions that need to be answered and (iii) the kinds of data that are preferred and most credible. The manager/ management team is also likely (iv) to want to make use of the results to make decisions about the direction the program should take to optimise success and (v) report the results to senior managers in the organization.

In the typical evaluation process, stakeholder groups do not share the same level of power. Power differentials influence

many, even all, aspects of an evaluation including: the terms of reference for the study; which stakeholders' interests an evaluation should serve and which it may ignore; the precise questions that it may pose and must answer; who should undertake it and when; what methodology or methodologies may be used; the nature of the data that may be collected; how the data must be analysed; how much may be spent and on what; the nature, style and length of the report; how the results should be presented; who may have access to the results; who may authorise the release of the findings and when; who may publish the findings and when; and how the results will be used. At each of these steps there is the potential for prejudiced decisions to be made that favour one group's self-interest and so fails to respect democratic practice.

While much has been written in the evaluation literature on the theory of evaluations and on specific cases, there is still no comprehensive and easy-to-use (mechanism) that can be used to hold organizations to the principles of effective evaluation.⁸⁵

In our experience, the body that controls the funds for the evaluation tends to exercise greatest power. It generally does so by deciding on the focus and scope of the evaluation, drawing up terms of reference, and contracting the evaluator(s). This body may be the donor agency who supplied the project funds or the executing agency itself. It is seldom, in our experience, the executing partners, or the third party beneficiaries.

⁸⁵ Source: Lempert, D.H. (2010). Why Governments and Non-Governmental policies and Projects fail despite "Evaluations": An Indicator to Measure whether Evaluation Systems incorporate the Rules of Good Governance. *Journal of MultiDisciplinary Evaluation*, Volume 6, Number 13. <http://www.jmde.com/> p. 58

Evaluation use: an example

In early 2005, the Papa Andina Regional Partnership Program, based at the International Potato Center in Peru, engaged one of the authors (Horton) to facilitate a process of reflection and evaluation over several months. Papa Andina was nearing the end of its third funding phase, and the initiative's main donor, the Swiss Agency for Development and Cooperation (SDC), needed an evaluation report. Papa Andina's members also wanted the evaluation to help them improve their work.

A utilization-focused evaluation approach was adopted. In initial meetings with Papa Andina's coordinator and a few members, several groups of key stakeholders were identified, including members of the Coordinating Unit, Strategic Partners, Steering Committee, CIP and SDC. Papa Andina's coordinator was careful to ensure that SDC, the Steering Committee and CIP were on board with the approach. There were initial questions and concerns about the amount of time the approach would demand of network members. However, when it was explained that evaluative activities would contribute to knowledge sharing, learning and program improvement, and would feed directly into planning for the next phase of the program, commitment to UFE was secured. Through discussions with key stakeholders, the key purposes of the evaluation were identified and the evaluation was subsequently designed to fulfil these purposes.

The evaluator and Papa Andina's Coordination Team jointly prepared an initial proposal for the evaluation, including timeline and budget. The Coordination Team then negotiated the evaluation proposal with CIP, SDC and the Steering Committee. The final proposal combined Horizontal

Evaluations, preparation of synthesis reports on major activities, a participatory evaluation workshop, and preparation of an evaluation report by two external evaluators. Efforts were made to combine the evaluation with activities and events already planned by the Program.

Several events including one horizontal evaluation workshop were organized to analyze and evaluate the major methodologies and activities developed by the program. Papa Andina's members were involved directly in data collection, analysis, interpretation, and reaching conclusions. These interactions between the evaluation team and Papa Andina's stakeholders allowed for the clarification of points, correction of errors, and dialogue concerning the main findings. Based on these exchanges, the evaluation report was finalized and formally submitted.

The materials prepared for the evaluation and the final evaluation report were disseminated in numerous ways. Highlights of the evaluation were published in the Papa Andina newsletter. Many of the materials prepared for the evaluation were incorporated into a 'Papa Andina Compendium.'

As members of Papa Andina were directly involved in the evaluation (especially during the evaluation workshops), many of them realized things during the evaluation process that they could put into practice straightaway. For example, a researcher in Ecuador realized during a horizontal evaluation that he needed to improve local participation in activities he was organizing. A researcher from Peru at the same workshop realized that he could improve the impact of his work by involving local government officials in planning his activities. During the final evaluation workshop, participants identified several areas for improving the work of Papa Andina in its next phase – most notably, the areas of gender,

policy influence, and evaluation. These areas were incorporated into recommendations in the final evaluation report, which provided the basis for planning Phase 4 of Papa Andina. After the evaluation, Papa Andina recruited a new team member to work on gender and evaluation.

Non-use of evaluation: an example⁸⁶

There is among those who undertake agricultural R&D a general apprehension that less than positive evaluation results at any level will prejudice their future funding. One of the authors (Mackay) in 1996 led a team responsible for assessing the results of the previous five years' efforts of the International Service for National Agricultural Research (ISNAR). He observed that after submitting the team's final report, senior managers appeared to be merely relieved that the evaluation made no damning criticisms of their organization or its programs. The fact that it contained many practical suggestions as to how to strengthen the organization and address its reported weaknesses appeared not to interest them at all. One of the most trenchant had arisen out of criticisms made by several of its clients that as a promoter of organizational change in R&D centers in the developing world, it had a responsibility to model the systems it was advocating. During the following 12 months that the team leader of the evaluation spent at ISNAR, the newly appointed Director General (DG) never once referred to or discussed the report or any of its findings or recommendations. It was as if the review had never been undertaken or

⁸⁶ Mackay, R. S. Debelo, T. Smutylo, J. Borges-Andrade, and C. Lusthaus. 1998. ISNAR's achievements, impacts and constraints: An assessment of organizational performance and institutional impact. ISNAR: The Hague. and Anderson et al. (2000). Impact of ISNAR 1997-2001, Report of an ISNAR Team.

had never been read by the new DG. After the following quinquennial review in 2002, ISNAR was closed.

Those who undertake agricultural R&D may find themselves in a conflict of interests when it comes to undertaking evaluations. On the one hand, they may have a genuine desire to know the outcomes and impacts of their efforts; on the other, they have to encourage those who provide them with funds to continue to do so. If their projects are shown to be less than successful, their viability as research organizations is threatened; if they insist on reporting only good news, they risk their credibility. A concerned member of a major donor to agricultural R&D, referring to the CGIAR, reported:

“A primary objective driving many studies was to demonstrate impact, to show donors that their investments in center research were well spent, and thereby to mobilize additional resources. Departing, often unconsciously, from the classic scientific method of hypothesis testing to move towards a demonstration mode, methodological problems became increasingly apparent: selection of successful cases for IA studies, inconsistent use of counterfactuals, overestimating benefit attribution to center activities, and restricting the dissemination of less favourable studies biased results and undermined their credibility and value. Donors and an increasing number of critics, also within the CGIAR, began to challenge the accuracy and representativeness of the exceptionally high published rates of return. As a result, both the resource mobilization and accountability goals of IA studies were often not achieved.”⁸⁷

The strategy of positive reporting backfired and triggered scepticism and ultimately a negative response from funders.

⁸⁷ Matlon, P. (2003). Preface to D. Horton and R. Mackay Agricultural Systems

The lesson appears to be that to fund evaluations only to dazzle and impress rather than to learn and improve one's own performance and understanding may turn out to be a poor investment in the end.

Evaluation misuse: examples

The majority of evaluations in “development” projects in many internal agencies follow a single formulaic result, guaranteed by an apparently manipulated (or subconsciously homogeneous) process to yield a typical conclusion: “The transfer of money or skills to ‘poor’ beneficiaries made them ‘richer’ or ‘better off’ than they were without the transfer. The project should be continued and expanded.”⁸⁸

A major challenge presented by agencies that fund evaluations is their lack of transparency regarding how results are made available to legitimate stakeholders. Those who commission evaluation studies are less likely to release or publish reports that find projects lacking in success than those, which support positive impact (IAASTD, 2009).

Getting it ‘right’?

The following example refers to an evaluation study commissioned by an international donor agency dedicated to helping developing countries build their science and technology capabilities to effectively address development challenges including agricultural R&D. The donor agency, who was also the executing agency, contracted six expert

⁸⁸ Lempert, D.H. (2010). Why Governments and Non-Governmental policies and Projects fail despite “Evaluations”: An Indicator to Measure whether Evaluation Systems incorporate the Rules of Good Governance. *Journal of MultiDisciplinary Evaluation*, Volume 6, Number 13. <http://www.jmde.com/>

evaluators to assess its capacity development efforts, over the previous ten years, in six disparate organizations to whose research projects it had contributed funds. One of the authors (Mackay) was asked to determine the results of the agency's efforts to develop the research capacity of an international agricultural research center and of its scientists. Over the decade, the agency had contributed more than four and a quarter million dollars to 12 of the center's research projects.

The evaluator's preliminary investigations indicated that both managers and scientists in the international research center believed that neither it nor its scientists should be the locus of the capacity development evaluation. The appropriate locus, they asserted, would be the regional research partners with whom it worked. The center justified its proposal by pointing out that in none of the 12 funded projects had the donor agency included capacity development for the center or its scientists as an objective. The evaluator suggested the change. However, the agency, while acknowledging that its funding had not explicitly included capacity development, claimed this as an implicit goal and insisted that the evaluator look for evidence of capacity development within the center despite its objections.

Accordingly, the evaluator developed a work plan that was approved by the donor agency and collected data using on and off-site document analysis and interviews. He analysed the data, prepared a draft report, sought written feedback on all but the conclusions from the scientists and managers interviewed, and revised the draft before submitting it to the agency in light of their feedback. It concluded that the agency's capacity development efforts had been relatively ineffective. Data suggested that the generally poor results were, at least in part, because the agency's capacity devel-

opment agenda was never made explicit to the target scientists. In essence, the agency used the steps in the competitive funding and reporting process as occasions to change the target scientists' research approach. A senior scientist with an international reputation, whose research projects had been funded in part by the agency for a decade, was quoted: "I knew of no capacity building intentions on the part of (the donor agency)." The implicit agenda was to move the scientists from a strictly disciplinary (e.g. plant breeding) approach to a multidisciplinary and more holistic developmental one. The agency's regional director acknowledged that scientists in the center would not perceive the funding agency as having a role in developing their capacity.

The report concluded that the agency's undisclosed agenda towards the research centre contradicted the definitions and approaches for capacity development that the agency supported as legitimate and effective. The report concluded that, in cases where scientists in receipt of agency funds had agreed to introduce a social perspective into their biological research, it was not possible to determine whether this was the result of conviction or of coercion. The evaluator provided data, which supported the latter while reporting that some scientists, using hindsight, appreciated the broader contextual importance of their biological research.

The agency's officer expressed dissatisfaction with the draft; it was boring, included too much detail and analysis, and the language was insufficiently diplomatic. The officer expressed concern about communicating negative results to the Board of Trustees and demanded a revised draft from the evaluator. The evaluator made stylistic changes to the draft and re-submitted. Abruptly, the agency terminated the evaluator's contract. Unilateral action, essentially the right to "fire without cause" is permitted within the conditions of

the contract this donor agency requires its contract evaluators to sign.

The agency subsequently employed a replacement evaluator, a qualified individual who had previously been an officer employed as a team leader by the agency for many years. The new evaluator undertook additional interviews and an expanded document review. The resulting report praised the agency's decade-long efforts to develop the capacity of the target research centre and its scientists. Positive comments and conclusions were gleaned from the previous evaluator's two draft reports, and expanded upon.

Government agencies, implementing organizations and businesses often lack adequate evaluation systems and in their absence routinely "thwart oversight regulations for transparency and accountability" in order to assure continuous streams of funding. Organizations on the receiving end of critical findings tend to "wave off" criticisms as subjectivity or an ill-informed perspective on the part of the evaluator. Lempert points to the absence of an adequate framework to determine if the evaluation policies of public and private organizations are in line with good governance. They may do no more than protect the status quo within the organization.

The manner in which organizations use evaluations for public relations purposes openly or covertly can also constitute a form of abuse at least from the perspective of the intended consumers of evaluation. The two primary purposes of the research evaluation the CGIAR has undertaken are described as "to obtain and justify funding support" and

“to support making decisions about the allocation of funds among alternatives”.⁸⁹.

Donors are becoming less confident of agricultural R&D’s capacity to address world hunger through genetic improvement alone and without donor guidance and have become less trusting of economists’ high-published rates of return.

The provision of evaluation results are suspect because their intention may be motivated as much by self-interest as the furtherance of knowledge and may be regarded as a form of abuse.

Challenges arising from within the evaluation community itself

In the absence of certification for practicing program evaluators, existing qualifications tend to be discipline-bound. Under such circumstances, those who practice evaluation may be limited in their approaches and methods to those of the discipline within which they have been trained. Where a specific disciplinary view of what constitutes legitimate evaluation predominates, this can have a detrimental effect on the growth of evaluation. As one example, the cadre of evaluation specialists within the CGIAR, a major consortium of international agricultural research centers, is made up predominantly of economists. It is not surprising, therefore, that a report on research evaluation requirements in the CGIAR, equates agricultural research evaluation with economic impact assessment. From the authors’ perspective, the two primary purposes of the research evaluation are “to obtain and justify funding

⁸⁹ Alston, J. M., Norton, G. W., Pardey, P. G. Science under scarcity: principles and practice for agricultural research evaluation and priority setting. CAB International.

support” and “to support making decisions about the allocation of funds among alternatives”. To accomplish these two purposes, the report explains that the CGIAR employs quantitative approaches to evaluation based on economic principles. It acknowledges that other approaches may be better suited to serve other purposes such as management decisions but regards these purposes as secondary and their methods inferior. From the perspective of these economists, monetizing results i.e. putting monetary value on all project outcomes and impacts is a *sine qua non*.

What is ‘Good Evaluation’?

The matter of what constitutes good evaluation and whom the appropriate judge might be is fraught with challenges.

One of the authors (Mackay) undertook an evaluation of a project to promote neglected and underutilized species (Andean grains and Asian millet) for the International Fund for Agricultural Development (IFAD). The project had ended a year prior to the call for the evaluation. The purpose of the assessment was to determine the extent to which project-initiated activities were being sustained beyond the funding period. The contract was for 23 consecutive days and the entire budget for all travel expenses and fees was under \$30,000.

A task force made up of members of the funding agency and the executing agency identified four questions for the evaluation to answer. The evaluator prepared an approach that would provide data sets credible to the principal information user (IFAD) to provide answers to the questions. The workplan was approved.

The questions and methods employed are described below.

- How appropriate was the project design and how conducive was it to accomplishing project goals? Methods employed included (i) a review of the initial design process, (ii) a content analysis of the project's logical framework, and (iii) a review of the project's management arrangements.
- What have been the project results and impacts and who are the primary and secondary users of these? Methods included (i) document reviews of planned project activities, outputs, impacts and beneficiaries as represented in the project proposal and (ii) a comparison of plans with annual reports to corroborate that planned activities had been conducted and had resulted in anticipated outputs. These produced three data sets: (a) an inventory of project results, (b) a matrix comparing work plans with achievements, and (c) an inventory of beneficiaries.
- How did the project bring about its results? Approaches to address this question included (i) scrutiny of implied program theory, (ii) program document review, (iii) interviews with regional coordination teams, regional partners, and beneficiaries to challenge project theory implied in the logical framework and in the results chains elicited from the regional coordinators, and (iv) direct observation in the field. These produced detailed case study patterns of plausible linkages between project activities and results observed during site visits. Hypotheses were developed, then data were collected and analysed, patterns were sought, checked, challenged, and refined during site visits and interviews with stakeholders.

- What has been learned during the project that could contribute to replication or scaling out? Methods included (i) gleaning and reviewing lessons learned from the annual national and regional documents, (ii) identifying and aggregating those that might relate to scaling out and then (iii) confirming the robustness of these lessons during confirmatory interviews with stakeholders at sites in both India and Bolivia.

The TORs also requested that the impact chain be identified showing the “relationships between activities, outputs, effect and impact, showing the sequence leading to the achievement of a given impact (either positive or negative)”. The impact chain was accomplished by using all of the data sets collected to identify technical and institutional causal mechanisms that were then used to confirm or disconfirm the results chains as described by the project’s managers and scientists.

The evaluator delivered his data and his findings and used them to facilitate a two-day participator workshop attended by key managers of the funding agency (IFAD) and key managers and scientists of both the executing agency (IPGRI) and its regional partners in Nepal, India (MS Swaminathan Research Foundation) and Bolivia (PROINPA) from whose projects much of the evaluation data had been collected during the evaluator’s field visits.

All of these parties reported that the workshop and the report had met their information needs in a useful way. The funding agency noted that the report allowed it to make decisions about future funding which it would not otherwise have found easy. Buoyed by the enthusiasm expressed by all stakeholders for the information delivered at the workshop and in the evaluation report, the executing agency subse-

quently submitted the report to its technical advisory office as an example of a quality evaluation. The advisory committee faulted the evaluation for failing to express impacts quantitatively.

From the perspective of the principal decision-makers, the study presented was a good evaluation. It provided them with both insightful and useful results including an identification of key elements in the impact chain. From the perspective of the technical advisory officer, the evaluation was deficient as an impact study in that it had not used an economic approach to assessing project impact. The credibility of the data collected and the analyses made in answering the principal information users' questions to the satisfaction of the principal information users was of little interest to the advisory committee. In essence, the two audiences were judging the evaluation from two separate perspectives.

Suggestions

In this section, we offer four suggestions for improving the use of evaluation in the agricultural R&D community.

1. *Employ evaluation as a tool to enhance learning and build knowledge in a cumulative way.*

This would have the effect of having those who commission and conduct evaluations question the value of 'producing still more and more of their one-off studies'⁹⁰ as opposed to generating 'analytic streams of evaluative knowledge'. These authors contend that it will be by conducting studies in such

⁹⁰ Mayne, J. and Rist, R. (2006). Studies Are Not Enough: The Necessary Transformation of Evaluation. *The Canadian Journal of Program Evaluation* Vol. 21 No. 3 Pages 93–120. Canadian Evaluation Society

a way as to accumulate bodies of knowledge and by synthesising that knowledge and integrating it into management practices and policies that evaluation will reach its potential for facilitating organizational learning and improvement.

2. *Allow the questions that stakeholders care about rather than disciplinary methodology drive evaluation.*

If heeded, this would have the effect of directing evaluators' attention to stakeholders' questions and the most appropriate/ adequate ways of answering these. Evaluations would be judged in terms of their relevance to and adequacy in meeting stakeholders' information needs. There would be potential for the ideological rift between opposing groups of evaluators to be bridged.

3. *Embrace the essential complexity and multidisciplinary nature of agricultural R&D.*

This would ensure that important contextual factors in R&D projects were given the attention they merit and prevent them from being ignored as externalities because they cannot be accommodated within a preferred model. It would involve experts from various disciplines working together to resolve complex questions and thereby more fully address the multiple factors that contribute to the success or failure of R&D projects.

4. *Accept that a legitimate function of evaluation studies can be to reduce uncertainty in the face of genuine complexity as opposed to provide uncontroversial solutions within a simulated simplicity.*

This would result in evaluators seeing the best approximate answer to complex questions as opposed to either ducking

such questions or pursuing more precise answers to simplified versions of the questions.

We have argued that agricultural R&D and its evaluation are important and complementary activities within a highly complex system. While an ecological view corresponds more closely to the complexities of the real world than does the laboratory view of reductionist science, nevertheless R&D based on both views are necessary. From an evaluation standpoint, the latter view constrains evaluation approaches to experimental methods and studies of economic impact and so admits only those questions to which such approaches can provide answers. The former broadens the permissible range of stakeholders' questions many of which do not permit simple answers expressed quantitatively. In order to answer them in ways that stakeholders find insightful, comprehensible, and useful, it draws on multiple methods from a wide range of disciplines.

If these views on R&D were accepted as complementary rather than adversarial, the entire world would be better served. Each perspective could bring its unique potential to improve the system as a whole. The agricultural R&D evaluation community could unite efforts and talk with one voice. Together they could refine existing methods and develop new, multi-disciplinary approaches to undertake evaluations capable of addressing the full range of questions. If we are to increase our understanding of what it takes to bring about successful international agricultural R&D before irreversible damage is done to our planet and those who live on it, questions must drive evaluation, not the reverse.

CHAPTER SEVEN

Changing Perceptions of the Environment and Environmental Evaluation

Ramon Perez-Gil

Grandma is sick. It took us many years to accept it, to grasp the long-term implications of this fact. Although over the years some people told us repeatedly about her symptoms, no one really wanted to believe she was actually sick. No one took seriously the sick condition of the long time provider of all our needs, or even to admit that historic roles needed to change. Grandma is now calling for our assistance. In retrospect, we should have noticed her symptoms: the coughing, the runny nose, and the loss of resilience, but we all tended to downplay the symptoms or blatantly overlook them. Today, as the cumulative effects of years of discomfort, aching, and abandonment are readily visible, the change in Grandma's overall performance is obvious –the high price of everybody's negligence. As the truth sinks in, we are finally truly sorry and concerned. Now the whole family is worried and some are taking an active role in trying to change the bleak prognosis that the crisis is only going to get worse.

Of course, this chapter is not about Grandma. It is about the environment, and the previous paragraph is relevant when

we read it again replacing the word Grandma for either one of these words, Nature or Planet. I cordially invite you to repeat the reading using these terms.

Any human activity, no matter if social, cultural or economic, takes place in a biophysical context and consequently interferes with the context to a lesser or greater degree. Every single human activity thus has an impact on the planet. Until recently, the generalized perception in most human societies was of an endless capacity of Mother Nature to keep providing us all of her goods and services and also all of her unlimited capacity to tolerate the harm infringed by an ever-increasing litany of destructive human activities. When people finally realized and fully understood that this was a false reality, they also began to understand the need to transform the quality and intensity of the relationships between humans and nature. However, this fact took centuries to be understood, as in the Grandma's allegory, many voices tried to pinpoint the evidence but the pace of enlightenment remained very slow.

This chapter examines the evolution of perceptions about the environment, particularly those that have occurred over the past 40 years. It has been an interesting educational journey in which assumptions have changed and attitudes have evolved. The history of the world's environmental enlightenment is a tale of resources, politics, economics, facts, feelings, prejudice and power. It is a story of systems, organizations and institutions evolving to fit improved understanding of reality.

Evaluation related to the environment has also evolved. Fifty years ago, before people perceived an environmental problem, there was little understanding of the need for evaluation. As the problem clarified and conceptions evolved, so did our notion of appropriate forms of evalua-

tion. When the problem is defined in terms of a lack of compliance, then compliance evaluation is in order. As understanding deepens and embraces complex intertwined human social systems, then deeper forms of evaluation are demanded.

Changing Consciousness and a Changing World Response

Many scientists contributed to enlarging our understanding of human-nature interactions, but among them Charles Darwin made a monumental contribution by incorporating biology into history to better understand the evolution of species. The input of other pioneers in ecology are worth highlighting, such as that of Stephen Forbes, and his Microcosmos concept, in which biological and physical processes interacted, or F. E. Clements who developed the concepts of succession and community, A.G. Tansley who introduced the notion of ecosystems and flows within them that was later consolidated by Eugene Odum. He simplified ecological analysis to states of development and mature states, the ecosystem with clear flows of energy and trophic levels became a standard concept. Later on, Pickett and White pinpointed the importance of disturbances in nature. The popular description of Rachel Carson's *Silent Spring* and the Club of Rome report *The Limits of Growth* (that explored different scenarios and stressed the choices open to society to reconcile economic progress within spatial environmental constraints), should be mentioned as true landmarks. They were indeed significant input that shaped societal response to the emerging realization that environment did matter. Not that environmental problems were new, some are (and obviously were then) very old, but indeed new problems

emerged and others were enlarged and threatened to become uncontrollable.

In the decades of the 1960s and 1970s, major evidence of a deteriorating environment were documented and brought forward, but mostly as isolated voices, but the frequency increased, and as we now know the turning point of the environmental crisis can be tracked back precisely to those times. The continuing onslaught on the environment has brought about negative consequences for humanity related to degradation of the planet's environmental conditions as well as an impoverishment of the biophysical and social systems that support our living conditions. This is now referred to as the environmental crisis, a compound problem whose most visible effects are climate change or global warming, the loss of biodiversity and the water crisis, all of which are symptoms of the critical situation in which we are living. Even though environmental destruction prompted reactions around the world, the snail's pace response of society to the environmental challenges is nerve-racking. For example, it took London until the mid 1970s to react to the air pollution documented at least 200 years before, in the mid-1600s.

It was only in the 1970s that the messages concerning environmental issues began to be finally heard and in a way dealt with, though most responses were still superficial. The rationale behind this limited awakening was perhaps the growing and evident destruction of the environment by people. The environment became a subject of research, though motivated by selfishness, as the deterioration of natural resources was affecting humans. The scientific community began to realize also the need to use knowledge from widely varied sources and disciplines, for the need to

find answers and sustain life on earth demanded creative and urgent approaches.

The voices expressing concern were eventually listened to in 1972 when the United Nations convened the UN Conference on the Human Environment, the first “environmental” World Summit in Stockholm, in which environmental matters were collectively brought to the world’s attention for the first time. It is fair to say that with the exception of a steadily growing number of environmental organizations worldwide, no one really had come up with a comprehensive proposal on how to deal with global environmental problems. Efforts were partial, scarce, local and well-intended but lacked information and above all were disconnected. Many were also skewed, as the first summit itself, towards the “human environment” short-sighted perception.

There were also huge vested interests in preserving the status quo. I remember well the time of the World Big Dams Congress, which was held in Mexico City in the 1970s. I tried at that congress to present the case of the Chicoasen Dam, a hydroelectric facility to be built in the state of Chiapas in southern Mexico, where the construction and all the associated effects, particularly the social and the environmental impacts, were huge and people such as myself were fighting to get more governmental and public attention to prevent and mitigate the myriad negative consequences. Another congress colleague pushed in similar terms the case of the Itaipu Dam on the border of Brazil and Paraguay. Both of us were kindly invited to reconsider our presentations for they did not fit in the program. All interventions were showing the great impact environment was having on the construction and other infrastructure. Clearly rain, storms, dampness, and even cold weather and snow were destroying the

hard work of men. The dominant perspective of that era was that human made structures were being damaged by the environment, so the environment had to be controlled.

As a result of the 1972 UN Conference on the Human Environment, a recommendation was put forward to establish a UN environmental organization and the United Nations Environmental Programme (UNEP) was created the same year by the General Assembly. The failure of the prevailing development scheme caused the international community to revisit their perceptions and attempt to make changes. In fact, the newly created organization was just one of many innovations. Changes in educational sciences led to the creation of environmental education as a new field.

In the beginning, attention was centered on issues such as natural resources conservation and preservation when it came to threatened or endangered flora or fauna. With the passage of time, other elements have been incorporated into environmental education, such as the technological dimensions and the social, cultural, political, and economic considerations that were lacking and are necessary to understand our relationship as a species with our environment.

The term environmental education was included in United Nations Science, Culture and Education Organization's (UNESCO) documents as early as 1965, but it was not until 1972 during the Stockholm summit when the concept was fully acknowledged and considered critically important to change society. No wonder with the onset of UNEP this concept was portrayed as a key action to protect our environment. The Stockholm summit was the catalyst for many initiatives, some of which only saw the light of day many years later. The so-called environmental debate was born.

In 1973, the Convention on International Trade in Endangered Species (CITES), one of the first international environmental agreements was adopted. Since then, a growing number of accords, agreements, protocols, conventions, treaties and the like, related in one way or another with environmental protection have been signed by a steadily growing list of countries and organizations. A few examples are worth mentioning to illustrate the nature of the perceived priorities in the global or regional environmental agenda. In 1975, the first UNEP brokered Regional Seas Agreement reached fruition; The Mediterranean Action Plan; the first intergovernmental conference on environmental education took place in 1977 in Tbilisi; in 1979 the Bonn Convention on Migratory Species (CMS) was signed.

In 1980, the International Union for the Conservation of Nature and its resources currently known as the World Conservation Union or simply IUCN, the most influential and leading environmental organization worldwide, launched its World Conservation Strategy, the first comprehensive document of a global scale dealing with environmental issues. It is fair to say that before this milestone (and naturally other events that we have mentioned also) the integral or holistic concepts of nature and the environment in general were not fully taken into consideration in decision making processes nor in monitoring and evaluation procedures.

In the early 1970s, the increase in human population together with oil spills in the oceans and carbon dioxide emissions seemed to be the biggest threats to human well being. By the beginning of the next decade, however, the perception of problems widened. The greenhouse effect was being discussed and also ocean pollution, deforestation, the loss of biodiversity and even acid rain, and by the end of the

decade other factors were incorporated in this perception but not just as threats to humans and their property or structures but to life on the planet and the planet itself.

That is why the GAIA hypothesis, although suggested back in 1965 by J.E Lovelock, gained popularity and relevance more than 10 years after. The hypothesis states that the temperature and composition of the Earth's surface are actively controlled by life on the planet. It suggests that if changes in the gas composition, temperature, or oxidation state of the Earth are caused by extraterrestrial, biological, geological, or other disturbances, life responds to these changes by modifying the abiotic environment through growth and metabolism.

During the 1980s, the perceptions and concerns widened even further to incorporate the severity of global climatic changes, the depletion of the ozone layer associated with CFCs, toxic residues including nuclear wastes and their final disposition, the loss of habitats, the pollution of water, both superficial and ground or fossil waters, the steadily increasing scarcity of freshwater, the intensification of environmental degradation and destruction, the shrinking of natural ecosystems worldwide, the waste of energy, the losses of soil, desertification, and marginalization. All these are among the examples of problems that comprise the complex environmental crisis as currently understood.

The world entered a phase of fine-tuning some of the existing international agreements by updating them. For example, the 1985 Vienna Convention for the Protection of the Ozone Layer was replaced two years later by the Montreal Protocol on Substances that Deplete the Ozone Layer. Indeed, the world could brag about some reduction of the indiscriminate and irrational use of toxic substances like DDT, lead, asbestos, dioxins and CFCs, as well as limited

achievements in the implementation of some of the international legal instruments to protect the environment and species. Regardless of the partial successes and the numerous efforts, the environmental crisis was still rampant and unmitigated -- presumably we were overlooking something.

In 1987, the Report of the United Nation's World Commission on Environment and Development: *Our Common Future*, also known as the Brundtland Report, alerted the world to the urgency of making progress toward economic development without depleting natural resources or harming the environment. Published by an international group of politicians, civil servants and experts on the environment and development, the report provided a key statement on sustainable development, defining it as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs." The debate was beginning to move beyond controlling particular commodities to a more holistic understanding of the interconnectedness of the world's ecosystems. The following year, 1988, still related to global changes in the environment, the Nobel Prize winner was the Intergovernmental Panel on Climate Change (IPCC).

Twenty years after the Stockholm summit, in 1992, the UN convened the Conference on Environment and Development (Earth Summit) in Rio de Janeiro, Brazil, an event in which Agenda 21, a blueprint for sustainable development, was launched, as well as other instruments like the Global Environmental Facility (GEF, nowadays called Global Environmental Fund) or the Convention on Biological Diversity (CDB) and the Private Sector's publication "Changing course", a global business perspective on sustainable development and the environment. This event was undoubtedly a turning point in global environmental

politics. Since Río there has been active promotion of sustainable development policies. A new legal international framework was launched through a number of multilateral environmental agreements (MEA), which include legal instruments that seek to establish more stringent regulations for social and economic actors hoping to limit and reverse the negative impacts of economic and technological processes on the environment.

The MEA also includes the Kyoto Climate Change agreements, the previously mentioned CMS and CDB conventions, the Convention against Desertification and Drought, and the Stockholm accord on organic persistent pollutants. Among these instruments, perhaps the most controversial have been those conventions and respective protocols or specific agreements related to climate change as well as those related to biodiversity, given their global character and the range of interests and conflicts to be deliberated//considered/pondered that each implied. The difficulties and challenges to internalize the ecological costs into mainstream economics were readily evident.

The fine tuning and debate continued. In 1995 UNEP launched its Global Programme of Action (GPA) to protect marine environments from land-based sources of pollution. In 1997, through what has been called the Nairobi Declaration, UNEP redefined and strengthened its role and mandate.

In 2000, the Conference of the Parties to the Convention on Biological Diversity adopted a supplementary agreement to the Convention known as the Cartagena Protocol on Biosafety. The Protocol seeks to protect biological diversity from the potential risks posed by living modified organisms resulting from modern biotechnology. In the same year, the first Global Ministerial Forum on the Environment called for

strengthened international environmental governance through the Malmö Declaration. In 2002, 20 years after Stockholm and 10 after Río, with great expectations surrounding it, the World Summit on Sustainable Development (also known as Río + 10) took place in Johannesburg, South Africa and shook governments and society in general with pessimistic scenarios for the planet's future.

In 2005, the Millennium Ecosystem Assessment highlighted the importance of ecosystems to human well-being and the extent of ecosystem decline. It underlined the key role of the natural environment in sustainable development, a leap forward again on human perception.

The effective results of the aforementioned efforts were dependent on and consequently reflect the prevailing understanding of the problems as to the way they seem to be affecting the planet and the life of humans on it, but also reveal the relative importance environmental issues had in each period. Inextricably linked to the relative importance of the problems is the level of resources (financial, material and human) devoted or allotted to deal with the problems. The greater the amount of resources devoted, the better the information gets collected, sorted and analyzed and in the long run the better the problem is understood and the better we are able to prevent, mitigate or deal with it. The efforts to perceive and understand environmental problems also made humanity revisit our relationship with nature, hence redefining the nature of environmental problems, their intensity and quality, and their relative importance in terms of the environmental agenda.

Evolving Perceptions of the Environmental Challenge

The different states of the progression of understanding nature and environmental problems and their causes correspond to different ways of describing the fundamental relation between humans and nature. This evolution in perception went through five paradigms (i.e. widely accepted perceptions of how reality is organised). The paradigms described herein are presented as sequential, which seems to correspond with the overall direction this evolution has taken. In practice, however, they overlap, occurring simultaneously in portions of society (i.e. sectors or countries, or even within a country) and in some contexts they may exhibit a different sequence.

Economic Pre-Eminence – “Man dominates nature”

This was the main paradigm in the industrial countries for almost 30 years following the World Wars. This paradigm asserts that nature exists as an instrument to benefit humanity, to be explored, exploited, manipulated, and modified, no matter how, as long as it renders an improvement in the material quality of human life. Nature corresponds at the same time to an unlimited source or offer of physical natural resources and as an endless and open receptacle of the by-products and wastes of the systems of production and consumption. In theory and in practice, economy and nature are separated and the economic processes of production and consumption occur in a totally closed system in which the only limiting factors are work and capital. The rest is shaped

by technological progress with its indisputable capacity to solve problems.

This approach generated a means of relation between human activity and nature; of a unilateral orientation, anthropocentric, where environmental damage, when even noted, was thought to be logical and easily repaired thanks to technology and its advances. The weakness of this approach emerged from the difference between vulnerability and ecological damage of tropical versus temperate ecosystems and from the differences between the types of environmental problems that each face. Until recently, the perception claimed that only the damage imposed to certain tropical areas was irreversible, but that the environmental problems of industrialized countries were entirely different and could be solved just by reducing pollution and the generation of wastes.

Furthermore, the amplification of commercial exchanges became the law and it slowly gained a global scale invading all territories and aspects of life. It is perhaps this over-economization of the world that induces the homogeneity and uniformity of production and consumption patterns and hence nourishes the simplification of landscapes and jeopardizes sustainability based in cultural and ecological diversity.

Ecological Pre-Eminence – “Back to mother nature”

Considered as the opposite paradigm from the one just described, back to nature even became a political movement and inspirational belief, an ethical system, an innovative value system reacting to the consequences of the dominant system (Deep Ecology). In contrast with the previously

described perception, humanity is placed in a subordinate position in relation to nature, this in turn supports its basic dogmas, the equality of all species, reduction of human population, bioregional autonomy (reduction of dependency, technological, cultural, economic and in trade, only within integrated regions with ecological common or shared characteristics), promotion of ecological and cultural diversity, a new orientation for economy not towards permanent growth but towards no growth, end of domination of technology, more use of native societies' technologies and resource management schemes and methods. Needed changes were indeed radical given the pre-eminence now suggested of ecological matters.

This reaction to Paradigm A is the other extreme. Its fragility is linked to its own non-viability given that it is not possible to expect that the world return to a life style that departs so dramatically from the current one, and for a great number of people, besides being impractical, it will also be undesirable.

Environmental Protection – “Damage control and preservation”

At the end of the 1960s, environmental problems in the industrialized countries such as pollution, habitat destruction, and loss of species placed greater attention to the environment and therefore the weakening of the prevailing economic paradigm. The strategy followed by the establishment was the “institutionalization” of the environment, of environmental impact studies manifested as a legal means to evaluate costs and benefits derived from pollution or habitat degradation. Governments created agencies to protect the environment, and to be responsible for the establishment of limits and quotas, standards, mitigation or corrective

mechanisms when said limits were passed, complemented with command and control instruments.

The acceptable, tolerated limits of pollution, for example, were determined by the short-term economic acceptance and viability of the enterprises, corporations or industries. No wonder most of them were arbitrary or lacked any technical or scientific support! Industrial cheating was generally justifiable because the correct ecological levels were not known. A lack of sound information prevailed. In industry, environmental management or concerns had as their main objective damage control. The limits were concentrated primarily in end-of-the-pipe measures instead of all the processes or all the steps in production.

The results of this approach in relation to the response of the enterprises are even less significant given that environmental management is seen as an additional cost and that the ecological gains or benefits will hardly ever translate into monetary returns for the corporation. In this paradigm, environmental problems are not yet assimilated as true limits, primarily given the omnipresent character of technology. Hence, the interaction between human activity and nature is kept unilateral and human centered, producing increasingly high negative impacts on nature.

Natural Resource Management – “Use and protect natural resources more responsibly”

Control and preservation also shifted, as more evidence was gathered and disclosed on the array of human-nature interactions. Humans and ecosystems and species could and should interact. In fact, in the real world we constantly interact. The anthropocentric view prevails. Natural resources ought to be managed so they can be continuously

used for humans' benefit. This rationale was justified as a wave of optimism seemed to blow among many circles reacting to the Club of Rome Report, supplanting their conclusions on scarcity. The main reason for the shift from paradigm C to this is associated with the increase and variety of ecological or environmental movements in many countries, both developing and underdeveloped. A basic topic of the Brundtland report, this approach has as its center axis the incorporation of all types of resources, biophysical, human, and financial, as well as infrastructure in the national accounting systems.

The shrinking availability or stock of natural resources was seen for the first time as a legitimate global concern. Pollution was seen not as a logical by-product of production but as a rather negative condition that is degrading our natural capital, the weather, and natural regulating processes. National parks, reserves and other protected areas were now valued as both reservoirs of natural resources and climate regulators. The concept of Biosphere Reserves, a management category of protected spaces where humans are also present, emerged under the umbrella of the Man and Biosphere Programme of UNESCO within this framework. The uniqueness of the concept and the explicit recognition of human presence and interaction was a departure from previous positions.

At the same time, the debt crisis of developing countries became more acute, stimulating a still more aggressive impact on natural resources because of an increased rate in extraction and destruction of natural resources. Debtor countries created this alternative to make payments more viable and to cope with the immediate needs posed by rapid growth in their respective populations.

The aforementioned factors fostered the strengthening of efforts, primarily by NGOs and some academic institutions, that led to the fine-tuning of techniques and methods of environmental monitoring and of information gathering and processing related to the depletion of natural resources and the overall impoverishment of the environment.

The management strategies associated with this paradigm, also dubbed “global efficiency”, include energy efficiency, natural resources conservation, ecological restoration, public health, ecosystem integrity monitoring, and the adoption of the polluter pays principle as a means to internalize the social and ecological costs of industrial pollution and to be a driver towards the wider use of cleaner technologies. In synthesis, the idea was to use market forces to create environmentally efficient management.

In the era of the politically correct greening of economies, nature shifts from being an object in the process of work to being coded in terms of capital. As a newly valued type of capital, or natural capital, it widens the ways and methods of economic valuation of nature.

Therefore, ancient forms of intensive natural resource exploitation coexist with the conservationist exploitation of nature. Biodiversity, however, is more than a multiplicity of life forms, but is also a reservoir of natural assets valued for its genetic importance, its economic worth for ecotourism and even its potential recognition as a carbon storage place.

Recent policies in biodiversity are not just dealing with the concern of the loss of species and their critical role in the planet’s stability. Biodiversity has revealed itself as a huge vault of genetic resources that is the prime matter of pharmaceutical and food industry corporations whose overall worth is beyond that of petroleum companies.

For the countries and particularly the people who live where this impressive array of ecosystems and species is located, the so-called megadiverse countries, this has not translated into any tangible return or reflection of their natural wealth. In fact, the opposite is true, for ancient local cultural perceptions towards these luxurious varieties of natural resources are oftentimes threatened or all together shaken or broken when only the economic value of such richness is considered. The biodiverse areas are not seen as they were in the past and this shift is shaping new forms of cultural, social, and economic strategies.

Ecodevelopment & Ecoefficiency – “Ecology Makes Economic Sense”

In these paradigms, environmental management is reoriented towards protection but with an innovative look, that of an open system, of a biophysical economy, where an open economy is thermodynamically built into the ecosystem. Part of the flow of biophysical resources (energy, materials, and ecological cycle processes) comes from the ecosystem towards the economy, whereas degraded energy and other by-products (wastes and pollution included) flow back to the ecosystem and consequently have an effect on the initial flow described. The underlying conception of ecodevelopment, its theory of development and environmental management, are based in the realization that humans and nature should not be taken as two dissociated entities, as most western philosophical approaches and governmental initiatives claimed for years. Humans are part of nature and human development takes place in nature; this is the biophysical context where all takes place.

One of the objectives of these complementary paradigms is to replace the polluter pays principle for the principle of

paying ahead for pollution prevention, through a restructuring of economics based on some ecological considerations. Ecodevelopment incorporates cultural and social equity concerns that were present in schools of thought close to the Deep Ecology philosophy, a movement aimed at blending biocentric and anthropocentric values into an ecocentric view.

Having evolved from the limitations of the former paradigms, the ecodevelopment and ecoefficiency paradigms seem much more suited to the present, for they imply more profound changes in theory and practice. The move to green economies is strong worldwide and through the principles of ecoefficiency many corporations are trying to integrate ecological uncertainties into economic models, into planning mechanisms including goal-setting, responsibility and benefit sharing, and means and methods selection (for production, consumption, relationships, growing or consolidation, etc.).

The ecodevelopment and ecoefficiency strengths have weakened. The underlying principles and foundations of some ecologist's proposals have been twisted and made to fit the mandate of the current economic rationale. The rhythms of natural resources exploitation and transformation have not slowed down, but on the contrary, they have intensified. Innovative strategies and interventions with nature have grown, creating new and unpredicted new environmental impacts and risks.

The economic and ecological imperatives that society is requesting have increased in number and complexity, yet the challenge of development and well-being in harmony with nature without destroying the environment remains.

Sustainable Development – “Satisfy current needs without jeopardizing the capacity of the next generation to satisfy their needs”

The world's concern towards the grave and diverse environmental problems the planet is facing found an alternative theory in the concept of sustainable development, coined years before and seldom used but becoming truly relevant when defined as in the Brundtland report of 1987. The famous definition recognizes that the term sustainable development is a binomial expression in which each part of the concept restrains or limits the other. Development was always perceived as a rat race in search of more – more of everything, speed, wealth, production, products, sales, etc. Whereas the word sustainable speaks of limits, the limits of nature, limits to provide if not given the time to refurbish or replenish, limits to tolerate damage, limits to process wastes, to cope with pollutants, limits to hold inhabitants even. The difference for many was a subtle one – changing the word sustained for sustainable seemed useless or meaningless – but how wrong they were. The word sustainable makes a great difference. Sustained implies it keeps a steady pace, increasing presumably without an inflection point forcing the curve to change. Sustainable development requires the promotion of patterns of consumption within environmentally sound limits. Social equity and justice are also involved in a new paradigm whose core is not to risk the natural systems that constitute the basis of our livelihood on earth, the atmosphere, the soil, water and the wealth of living organisms. Sustainability implies harmonizing different sides of human development, such as economy, society, nature, culture, and technology, where the environmental dimension is a cross cutting element in the process of development.

Today the current relationship of humans and nature and its perception can be seen in the discussions around the means to launch and implement sustainable development, and in the policy proposals and instruments and concepts of environmental management.

An Emerging Paradigm Perhaps?

In time, the environment as a whole began to be more properly considered, accounted for, and evaluated. This new conscience incorporates the full recognition of nature and all its components as active players. For many years, environmental quality remained as an indirect measure of the amount of pollution or crowding, but this seems to be largely a suburban middle-class ethic imposed by our western culture. In reality, many factors bear more heavily on the quality of life, such as watershed “healthiness”, green belts, ready access to natural open spaces, biodiversity in natural ecosystems, all the way through contemporary environmental services (pollination, water catchment & air cleansing, carbon sequestration). This fuller conception seems to have penetrated the environmental quality concept just very recently. The emerging scientific and pragmatic field in environmental management seeks to balance the demand of earth’s natural resources with the capacity of the natural environment to cope, in a sustainable fashion, with those demands. A drive to recognise intergenerational rights is a sensible and wise alternative to prevailing modes of development and even the globalization of an environmental solidarity.

As said before, the above mentioned paradigms are not necessarily mutually exclusive. They might coexist in time and space.

Environmental Tools and Instruments

Perceptions on the environment have changed over the years, as have goals in relation to environment or nature as whole. The reasons are many: major political or organizational events, institutional changes, world events, social pressure, or a simple shift in priorities can translate into a change of goals, and of course, leaders change goals at their will.

Everyone that sets a goal has to have some way to try to measure or assess progress towards reaching it. In the case of environmental goals, perceptions have varied in close relation with the shifts in paradigms described in the previous section. In the same way, approaches, methods, tools and instruments to assess or evaluate both progress towards goal completion and the status of the environment have evolved. These are of varied nature and may be preventive, corrective, remedial or proactive depending on the phase or stage in which they are put in place. The most important tools are briefly presented here.

1. Environmental Impact Assessment (EIA)

The evaluation of impact on the environment took several shapes and avenues. Road construction, urban development, big dams building, pollution control in industries, transportation, power generation, agricultural growth, etc. were challenged by the need to consider evaluating their impact on the environment. The Environmental Impact Assessment or EIA is one of the earliest tools for this purpose. It is widely accepted and most commonly used. The origins of the EIA, as a formal and institutionalized activity, has to do with the promulgation in the USA of the National Environmental Policy Act (NEPA) in 1969 and its disclosure and

eventual adoption in many other countries after the Stockholm summit of 1972. Since then, the EIA became widely known and it is perhaps the single most used and widely adopted environmental management tool. It has become part of the environmental policy in many countries. By incorporating the analysis of social, physical, biological, and even economic impacts, the EIA's strength goes beyond its quantitative aspects to its explicit identification of the damages and costs caused to the environment and society by destructive processes and agents.

2. Environmental Monitoring Programmes (EMP)

Considered as an essential instrument for any environmental management system, environmental monitoring implies the systematic follow up of the temporal and spatial variation of several environmental parameters, of which selection and interpretation of data is a part.

Its importance lies in the fact that the EMP enables the constant evaluation of the environmental management system as a whole by focusing on specific issues to attend to and solve, related to wastes, potential problems, costs or inefficiencies and even compliance with norms, standards, social or governmental demands and needs. The efficiency of this instrument is dependent upon the proper selection of the environmental indicators to be used, of the intelligent selection of the sampling points, control stations, period, frequency, registry, and proper analysis of data.

3. Environmental Audit & Environmental Systems Verification

Together with the EIA, the environmental audit has become one of the most commonly used tools particularly by the industrial sector. Currently its use is primarily of a voluntary

nature in Mexico, Canada, the USA, and Europe. The European Community defines it as a tool that comprises a documented, periodic, and objective systemic evaluation of an organization's performance, its managerial system and of the equipment designated for environmental protection. Its main objectives are easing the control and management of environmental practices and to evaluate its compliance with the prevailing environmental legislation. This approach has had a very interesting spin off in the form of the launching of a variety of economic instruments to encourage the industrial sector to do more on the environmental agenda and to go beyond mere compliance. These economic instruments include incentives like tax cuts, eco-labelling, and certificates of compliance that translate into a better corporate image.

4. Risk Analysis

An instrument commonly used in combination with the EIA consists of the identification of elements and situations of a given activity or of a given product that imply risks to the physical environment and to human's and other organism's health. A risk analysis includes: a) the identification and classification of the hazardous events through inspections, research, questionnaires, etc; b) the determination of the frequency of occurrence through probability calculation; c) analysis of the effects and collateral damage associated with the events through mathematic modeling, and d) definition of control and mitigation techniques and measures.

5. "Due Diligence"

Due diligence is an approach usually related to corporate buy-outs, mergers and similar financial operations between corporations but can also apply to other acquisitions and most recently, due diligence is being used as a tool for

environmental insurance. It comprises research activities that are intended to identify potential environmental obligations and costs, the so-called “environmental passives” of a previous owner. As part of this activity one studies the environmental history of the company or site, plus conducts inspections, sampling, and laboratory analyses if needed.

6. Environmental Recovery Programs

This constitutes an environmental planning and management instrument which is applied in the initial stages of a given project to influence the technical orientation of the project should potential problems and impacts be detected in advance. In particular, when processes that destroy or negatively affect environmental attributes are involved, a recovery program that restores the environment to its previous unaltered status must be part of the planning.

Realistically, once a site is altered, there are no human activities that can fully restore it to its previous condition. Most efforts, even those that are well intended and even science based only mimic in a very limited fashion the works of nature. Nevertheless, every recovery plan attempts to do so.

7. Environmental Emergency Measures Programs

Developed as a means to complement the risk analysis mentioned earlier, these programs involve the articulation of a series of actions aimed primarily at dealing with the potential emergencies. A comprehensive environmental emergency measures program must involve all relevant areas within the organization from its inception. It must foresee the ways and degrees to which each area ought to participate in the event of an emergency. The program must clearly state at least the intervention sequence to assure

efficiency and the highest degree of control possible in such an extraordinary event, including a communications strategy. Its scope and success is greater if it also includes preventive measures, training, and awareness plans on risk prevention and on emergency measures.

8. Communication Programs

Communication programs are seen as the most important complements of any environmental program, tool, system, or management scheme. They are well accepted by corporations and the public yet frequently not well understood for they are often taken merely as public relations exercises. Such programs seek to inform society on the activities and endeavours performed by corporations or governments in the widest environmental arena, but at the same time gather people's opinions, perceptions, and priorities.

9. Integral Accounting and Footprint Assessment

Several economic indicators are used worldwide to compare economies, countries, their relative wellbeing or presumably their level of development. Most of them are well known and make the bulk of regular publications produced by entities like the World Bank, the International Monetary Fund, United Nation's agencies or even the so called Think Tanks. These constitute the System of National Accounts, which is a conceptual framework **Error! Bookmark not defined.** that sets the international statistical standard for the measurement of a market economy. The System consists of an integrated set of macroeconomic accounts, balance sheets and tables based on internationally agreed concepts, definitions, classifications and accounting rules. Together, these principles provide a comprehensive accounting framework within which economic data can be compiled and presented in a format that is designed for comparative

purposes of economic analysis, decision taking, and policy-making. Being the most comprehensive macroeconomic standard, it also serves as the main reference point for statistical standards of related statistics such as balance of payments. Gross net income, Gross Domestic Product, industrial and agricultural production, capital investments, Consumer Price Indices, exports and imports, and even debt are among these macro indicators, indexes and parameters.

In the mid eighties, the realization that the traditional economic indicators were skewed and hence useless to reflect a more comprehensive reality (i.e. one that considered at least an array of social, cultural, ecological, and economic aspects) hit some governments and other entities. New tools emerged to try to cope with this need. One is the National Integral Accounting Systems, also called satellite accounting or simply environmental accounting system, which actually evolved from the System of National Accounts. The innovation came when governments realized that the costs derived from natural resources depletion and costs associated with negative environmental impact needed to be considered and in fact deducted from the accounting. As an example, overexploitation of forestry resources was always added as a positive in the country's overall balance. With the integral accounting system, such impact on the country's assets is instead deducted from the overall balance.

Another innovative tool that has already given birth to variations is the Ecological Footprint developed by William Rees and Mathis Wackernagel in the early 90's and its recent derivations like the Hydric Footprint and Carbon Footprint and equivalent applications. It has been portrayed as the world's premier measure of humanity's demand on nature. It measures how much land and water area a human

population requires to produce the resources it consumes and to absorb its wastes, using prevailing technology. The underlying rationale is that all the direct and indirect impacts of the current pattern of consumption of a given country is accounted for and summarized; hence one can compare the intensity of the footprint left by said country on the planet.

10. Organizational Assessment

As was the case with other types of organizations, organizational assessment of organizations dealing with the environment followed project and program evaluation. Beginning in the 1990s, the fact that organizations both influence and are influenced by the environment was understood in a wider way, and some development agencies that supported environmental NGOs, for example, began to consider ways of assessing organizational performance. Indeed, Universalialia's IOA framework was used to evaluate partnerships intended to strengthen partners working on the environment in the South. A decade later, Universalialia worked closely with IUCN to build that organization's evaluation capacities, and organizational assessment became a key component.

The Paradigms and the Toolbox

As the paradigms evolved, so too did the focus of enterprises and governments. Both the private sector and governments reacted to the pressure exerted by different segments of society for more responsible environmental engagement. The demands varied depending on the level of development of the countries and in relation to the nature of the issues and the level of environmental consciousness of

society. NGOs occupied the central arena in the circus and, as in a circus, they performed in various ways – as spokespersons of societies affected by certain environmental problems, as proactive consumers in need of cleaner or environmentally sound products, as information gatherers, managers and brokers, as think tanks, and some even as allies to certain interests acting as smokescreens distracting society and preventing a clear view of the issues at stake.

Once the pressure channels were properly established and acknowledged, even sometimes co-opted, governmental authorities reacted, issuing regulations against pollution and offering incentives to induce corrective and preventive measures aimed at less environmental degradation. The intensity and frequency of demands is less in underdeveloped countries, however, pressure may come from abroad, perhaps related to exports, or through foreign NGOs that pressure for the adoption of new policies. Change can also result when competition between corporations leads to an advantage for the one with more stringent environmental standards.

The behaviour of the business sector is still quite heterogeneous; being environmentally sound or at least environmentally conscious is not a rule, not even when talking about enterprises present in many countries at one time. The leading 16 corporations in mining, manufacturing, services and technology acting in countries like Canada, Denmark, France, Germany, and the UK, stated in a survey that the two major drives that originally determined their changes towards the environment were legislation and technological improvement. These were followed by the pressure of NGOs, clients and even employees, new business opportunities, quality control systems, accidents, and new global corporate policies.

One interesting example from the soft drink bottling industry is illustrative. A bottling plant in Mexico holds the world record of needing only 1.3 litres of clean water for every litre of the cola they produce; however, the global average is 3 litres of water for every litre produced and some countries use 6 litres of clean fresh water for each litre produced. The technology is there but not necessarily the willingness to put it to work and pay the bills of the required investment, regardless of its high return.

The main determining factor, therefore, is not yet environmental awareness; the key delineators of profit are in reality still the factors that decide the changes in corporate behaviour. Public opinion is important and its importance is growing steadily, but it remains secondary to profitability.

As can be seen in table 1 below, as time went by, more tools became available. The fact that not a single approach could provide all the answers made governments, universities, NGOs and corporations use more than one tool at a time. Arrows in the table refer to the actual usage of the tool in a given period. The relevance of the use of any tool can be argued for any given period, but the fact is that not all the tools were available during this evolution. For example, the Ecological Footprint concept and methods was first published in 1992, whereas EIA became popular in the early 1970s.

CORRELATION OF TOOLS AND INSTRUMENTS WITH EVOLVING PERCEPTIONS AND PARADIGMS

EVALUATION TOOLS AND INSTRUMENTS	PERCEPTIONS AND PARADIGMS → → → → TIME → → → →						
	ECONOMICS PREMINENCE	ECOLOGY (DEEP) PREMINENCE	ENVIRONMENTAL PROTECTION	NATURAL RESOURCES MANAGEMENT	ECO-EFFICIENCY & ECODEVELOPMENT	SUSTAINABLE DEVELOPMENT	NEW IN THE MAKING EMERGING
ENVIRONMENTAL IMPACT ASSESSMENT (EIA)			✓	✓	✓		
ENVIRONMENTAL MONITORING PROGRAM/ISO 14000		✓		✓	✓	✓	✓
ENVIRONMENTAL AUDITS/ECONOMIC INSTRUMENTS			✓	✓	✓	✓	✓
RISK ANALYSIS		✓			✓	✓	✓
ENVIRONMENTAL PASSIVES/DUE DILIGENCE						✓	✓
ENVIRONMENTAL RECOVERY PLANS	✓	✓		✓	✓		✓
ENVIRONMENTAL CONTINGENCY PLANS			✓		✓		
COMMUNICATIONS PR PROGRAMME			✓	✓	✓	✓	✓
INTEGRAL ACCOUNTING SYSTEM/ECOLOGICAL FOOTPRINT				✓	✓	✓	✓
ORGANIZATIONAL INSTITUTIONAL ASSESSMENT (OIA)	✓		✓		✓	✓	✓

During the economic prominence paradigm era an equivalent to environmental recovery plans and perhaps the evaluation of the organizations performance through a sort of organizational assessment were the only tools in widespread use to deal with environmental issues. As Deep Ecology became pre-eminent, natural elements were monitored to prevent potential damage to them, and the reactive tools of the era were the recovery plans and a variation of what we now call risk analysis. By the time the need to protect the environment became clearly understood and in a way built into plans and programs, of governments and corporations alike, the EIA and Environmental Audits became the key tools in use. Industries and governments wanting to show society that something was being done on the matter gave some attention to the first contingency plans

and the first tailor-made media interventions on environmental matters. The natural resource management era saw the growth of the ISO 14000 series on environmental matters, the birth of new approaches on national accounting systems and more maturity in other existing tools and instruments. The era where corporations embraced the notion of the combined ecological and economic efficiencies, still current in many places and in many sectors, has the broadest array of tools and instruments to ponder the impacts on the environment and to deal with environmental issues, beyond compliance or even corporate social responsibility. Perhaps the only one missing is a tool that includes a longer time frame that can be expanded to generations in fact, the accounting of cumulative effects through environmental passive analysis. The current paradigm of sustainable development builds environment considerations into the planning stages and therefore the contingencies and the need to compensate unforeseen negative impacts are reduced to a minimum. This is the theory; in reality, the tools and instruments are all still being used extensively as environmental matters have finally won their place on global and local agendas.

Final Remarks

Governments, corporations, civil organizations and the public in the most varied cultures and countries are gradually realizing that the environment matters, that it can no longer be overlooked, and that our future survival is at stake. Even though environmental consciousness seems to have emerged almost 50 years ago at the end of the 1960s, not until the last 10 years, perhaps as a spin off effect of the Rio de Janeiro summit of 1992, have the debates, concepts

and policies surrounding sustainable development and the environment become really understood by the public. Environmentalists have now placed terms and concepts previously reserved for scientific and academic circles into common language and public consciousness. To be sure, terms and concepts are oftentimes involuntarily and voluntarily confused – prevention, risk, environmental audit, restoration, environmental monitoring, impact, rehabilitation, mitigation, ecosystem, biodiversity. Such concepts might not necessarily be properly understood, but they are used and some positive results are harvested from this involuntary puzzlement.

The popularity of climate change has led to a Nobel Prize and the millions of dollars spent either to disclose its facts and implications or to sponsor fake controversies fuelled by commercial or national interests has had the double effect of, in some instances, pushing other environmental problems higher on the agenda. In other cases simply the opposite, casting a shadow on top of other equally important problems or symptoms. Indeed, some of today's environmental debates mask the true underlying causes of the ecological crisis. So, amongst the hot discussions on global warming, the entropic degradation that all economic activity entails under the prevailing economic rationale is neglected.

The geopolitics of sustainable development looks with optimism to the eventual resolution of the contradictions between economy and ecology when proposing the conversion of biodiversity into collectors of greenhouse gases (primarily carbon dioxide), this way freeing the industrial countries from being guilty of exceeding their pollution quotas, while inducing the ecological conversion of the third world countries. Through this mechanism, however, the risk is that the marketing of nature will only deepen the gaps

between rich and poor countries. The new globalization justifies the comparative advantages of the more industrialised and polluting countries as well as those of the poor countries that can both revalue their capacity to absorb the excesses emissions of rich countries (carbon sinks) and offer genetic resources and ecoturistic attractions derived from their biodiversity reserves.

Even though the environmental problems are so evident and their effects are jeopardizing the whole of humanity, like in the Grandma's allegory with which I began, not enough people seem to either care or see. The important issue is that the ecological conscience has been awakened. This includes all of us dealing with intellectual or academic activities, as well as those in governments and the private sector, leaders in every field. The number of publications, initiatives, organizations and fora are growing steadily, showing that we are indeed on a new path. History must be revisited to ponder the actual weight and role of environmental issues and to better understand the magnitude of the evolution of perceptions towards nature.

The future paradigm needs to shatter the reluctance for change and foster the end of the generalized political, cultural, and societal paralysis. The immobility of politicians needs to be replaced by institutional and organizational transformations and effective cooperation leading to higher standards in all countries, no matter whether they are rich or poor, developed or under developed, crowded or less populated, muslim or catholic, pink or yellow, black or white.

CHAPTER EIGHT

Evaluating ICT: A New Dress for Old Questions

Ricardo Gomez and Shaun Pather

The interest of researchers and managers in evaluating the effectiveness of Information and Communication Technologies (ICT) has been evolving since the first commercial deployment of computers. The search for appropriate ways to assess the benefits and impact of ICT, which are elusive and frequently intangible, has generated much debate in the field of evaluation, and has been compared to the search for the Holy Grail. In the late 1990s, the advent of the internet, which escalated the deployment of ICT beyond the traditional business world into mainstream society, opened up opportunities to use ICT in support of international and community development activities. This new dimension of ICT in support of development (also called ICTD or ICT4D) added a new layer of challenges to the already elusive subject of ICT evaluation. This chapter discusses the following question: how can we assess the direct and indirect benefits of the use of information and communication technologies in the context of international and community development efforts?

The application of ICTs as a developmental tool has progressed through various phases over time. Thirty years ago, sceptics viewed the use of ICT for development as a needless luxury: “How can we prioritize computers when people need water and roads?” was a common response by critics in the late 1980s. Twenty years ago, ICTs became a panacea that was euphorically believed to solve every development problem: “ICTs will bring about a new world of electronic democracy, prosperity, and education for all”, was an easy euphoric claim as we approached the end of the millennium. Ten years ago, however, much cynicism has followed the euphoria, fuelled in part by the many anecdotal stories of failure and success of ICT experiences in development that characterized much of the ICTD environment during the first few years of the new millennium. Despite widespread investment in ICT by governments and donor agencies, questions remain as to what are the impacts, if any, of ICT in development.

Today, a handful of researchers have started to ask hard questions about the evaluation of ICT impacts, wondering whether they do exist, and whether there is a way to reliably measure them, for example Heeks and TASCHA⁹¹. In recent years, nonetheless, ICT and development initiatives are being either abandoned, mainstreamed into broader sector development themes (environment, agriculture, gender, education, etc), or “gadgetized”: ICTD is becoming a lab where numerous technology applications, devices and gadgets are developed, hoping they will be adopted and

⁹¹ Heeks, R. 2009a. The ICT4D 2.0 Manifesto: Where Next for ICTs and International Development? Working paper series. Development Informatics Group, Institute for Development Policy and Management. Manchester. and TASCHA. 2010. Global Impact Study of Public Access to Information & Communication Technologies. Technology & Social Change Group, University of Washington. <http://www.globalimpactstudy.org>

used by people working in international development contexts.

In reflecting at the ICTD evaluation landscape, we suggest the ICTD evaluation field may be too narrowly focused on measuring the tangible and quantifiable economic benefits of ICT for development, and as a result we may have been “barking up the wrong tree”: Have we been so concerned with evaluating the tangible social and economic impacts that we have neglected other intangible impacts, which may be equally or more important for human development than the tangible and quantifiable ones? We argue that in the early days of ICTD we were focused on evaluating the outputs (counting number of computers, number of users, etc), and eventually started to look at other tangible outcomes related to economic growth: business opportunities, income generation, new markets, etc. This is convergent with the modernization paradigm of development as transfer of technology for economic growth. To elaborate this idea, we explore the experiences of ICT evaluation in the business world, which has a longer history of ICT evaluation, and draw some parallels, which inform how the goalposts in ICTD evaluation could be shifted. We conclude this chapter by suggesting a fundamental shift in the way we think about ICTD evaluation. We point the way towards a mindset which focuses on both tangible and intangible contributions of ICT to human development.

The Promises of ICT

One of the most prominent early proponents of the benefits of ICT to human development was then Vice-President Al Gore, who professed that "...we will derive robust and sustainable economic progress, strong democracies, better

solutions to global and local environmental challenges, improved health care and ultimately, a greater sense of stewardship of our small planet... [ICT] will help educate our children... it will be a metaphor for democracy itself..."⁹²

Two decades ago, statements such as this gave both the developed and developing world great hopes based on the potential of modern ICTs to assist an acceleration of development efforts and impacts. This sort of rhetoric was in line with the classic view of development as a 'fast track' to modernization, and of technology as an inevitable driver of social change ('technological determinism'). These views of development and technology were coupled with an implicit threat to "get wired, or else", based on the idea that "although the costs of using ICTs to build national information infrastructures, which can contribute to innovative 'knowledge societies' are high, the costs of not doing so are likely to be much higher."⁹³ However, just as the technology world was jolted into reality when the e-bubble burst in early 2000,⁹⁴ and the business world was shaken by the financial meltdown of 2008, the development sector has seen a slow melting of the dream of technology bringing about accelerated development, wealth and opportunity to the majority of the world's poor.

With the dawn of the new millennium, and notably the two-phased World Summit on Information Society (WSIS)

⁹² Gore, A. 1994. Address to the International Telecommunication Union, 21 March 1994.

⁹³ Mansell, R., & Wehn, U. (eds) 1998. *Knowledge Societies: Information technology for sustainable development*. United Nations Commission on Science and Technology for Development. Oxford University Press, New York.

⁹⁴ Remenyi, D., Grant, K., & Pather, S. 2004. It was a shock when Boo went under: the legacy of the e-bubble: lessons for managers. *Journal of General Management*, 29(3):24-36.

shortly thereafter (Geneva, 2003 and Tunis, 2005) we have witnessed a proliferation of research output in ICTD, supported by research agencies, non-profit organizations and academics. British researcher Richard Heeks⁹⁵ suggests that hundreds of millions of US dollars are invested each year in ICTD projects, and that the ICTD research area is growing significantly faster than other cognate areas. Furthermore, he posits that the ICTD outputs to date reflect: (i) a bias to action and not a bias to knowledge, (ii) a preference for what is narrowly descriptive and (iii) a field that is not analytical enough. Others suggest that shortcomings of this research area include a lack of theory, conceptual definition, interdisciplinary approach, qualitative research and longitudinal research⁹⁶.

After 30 years, the field of evaluation of ICTD is still maturing. In the post hype era of the millennium, there is an opportunity to refocus evaluation by moving beyond the modernization paradigm, and by closer introspection of intangible benefits. To inspire this path, we will first discuss the experience of ICT evaluation in the business environment, following which we present views regarding alternatives to the modernization approach to development.

⁹⁵ Heeks, R. 2009b. Worldwide Expenditure on ICT4D. ICTs for Development Talking about information and communication technologies and socio-economic development, Blog, <http://ict4dblog.wordpress.com/2009/04/06/worldwide-expenditure-on-ict4d/>

⁹⁶ van Dijk, J.A.G.M. 2006. Digital divide research, achievements and shortcomings. *Poetics*, 34: 34 (2006) 221–235.

Evaluating ICT - Lessons from the Business Environment

The challenges of ICT evaluation are not entirely new. For many years, since the first deployment of computers in business, researchers and practitioners have also grappled with evaluation of its impact. Just as we currently recognize the intrinsic connection between ICT and development, in the business sector there has long been the idea that the implementation of ICT is indispensable to the provision of effective organizational services. As a result, the implementation and management of ICT has presented both major opportunities and challenges to businesses. Amongst these challenges, the increased complexity of ICTs combined with the uncertainty and unpredictability associated with its benefits and costs, pointed [researchers] to the development of sound evaluation methods which offer companies a deeper insight into the impact of their ICT investment⁹⁷.

The business benefits of ICT gained much attention about twenty years ago, when economist and Nobel Laureate Robert Solow characterised the computer age by saying that “we see computers everywhere except in the productivity statistics”⁹⁸. This anomaly became known as the productivity paradox of information technology, and there were various reasons offered to explain this paradox, such as “deficiencies in [the] measurement and methodological toolkit” and the

⁹⁷ Irani, Z., Love, P.E.D. & Zairi, M. 2000. Information systems evaluation: mini-track introduction. In Chung, M. (ed.). Proceedings of the 6th Americas Conference on Information Systems, Long Beach, California, 10-13 August. Long Beach, CA: California State University: 1073-1075.

⁹⁸ Solow, R.M. 1987. We'd better watch out. *New York Times Book Review* 36, July 12.

“mis-measurement of outputs and inputs”⁹⁹. Other researchers before Solow had explored the potential business benefits of ICT as early as forty years ago (e.g. Boyd & Carson, 1963; Gallagher, 1974; Lucas, 1973¹⁰⁰). These early studies were concerned with whether technology was being effectively used, and explained this effectiveness in a variety of ways ranging from relatively simple accounting measures to complex multi-dimensional balanced score-card type metrics (Bannister, Berghout, Griffiths & Remenyi, 2006¹⁰¹); they also explored the use of surrogate measures such as user satisfaction, service quality, individual and organizational impact (DeLone & McLean, 1992¹⁰²; Lomerson & Tuten, 2005¹⁰³; Seddon, Staples, Patnayakuni & Bowtell, 1999¹⁰⁴; Whyte, Bytheway, & Edwards, 1997¹⁰⁵). There are three phases that describe how business interest in ICT evaluation has evolved over the years. These can be mapped against three distinct eras of ICT deployment in business.

⁹⁹ Brynjolfsson, E. 1993. The productivity paradox of information technology. *Communications of the ACM*, 36(12):66-76.

¹⁰⁰ Boyd, D.F. & Carson, H.H.J. 1963. Economic evaluation of management information systems. *IBM Systems Journal*, 2:2-23.a; Gallagher, C.A. 1974. Perceptions of the value of a management information system. *Academy of Management Journal*, 17:46-55; and, Lucas, H.C. 1973. User reactions and the management of information services. *Management Informatics*, 2(4):165-172.

¹⁰¹ Bannister, F., Berghout, E., Griffiths, P. & Remenyi, D. (2006) Tracing the eclectic (or maybe even chaotic) nature of ICT evaluation. In Remenyi, D. & Brown, A. (eds). *Proceedings of the 13th European Conference on Information Technology Evaluation*, Genoa, Italy, 28-29 September. Reading: Academic Conferences: 41-51.

¹⁰² DeLone, W.H. & McLean, E.R. 1992. Information systems success: the quest for the dependent variable. *Information Systems Research*, 3(1):60-95.

¹⁰³ Lomerson, W.L. & Tuten, P.M. 2005. Examining evaluation across the IT value chain. *Proceedings of the 8th Annual Conference of the Southern Association of Information Systems*, Savannah, Georgia, 25-26 February. s.l.: s.n.: 124-129.

¹⁰⁴ Seddon, P.B., Staples, S., Patnayakuni, R. & Bowtell, M. 1999. Dimensions of information systems success. *Communications of the Association for Information Systems*, 2(article 20):1-60.

¹⁰⁵ Whyte, G., Bytheway, A. & Edwards, C. 1997. Understanding user perceptions of information systems success. *Strategic Information Systems*, 6:35-68.

Laudon and Laudon¹⁰⁶ describe the evolution of these eras: “In the 1950s the effects of IS [Information Systems, or Information Technology] on organizations brought about merely technical changes, only serving to automate clerical procedures. During the 1960s and 1970s IS had an impact on managerial control, and from the 1980s onwards IS impacted upon core institutional activities such as products, markets, suppliers and customers.”

Zuboff¹⁰⁷ labelled these phases as Automate, Informate and Transformate: The Automate phase focused on measurement of technical aspects of IT; the Informate phase shifted towards evaluating the measurement of IT production or IT project management; the Transformate phase focused on measurement of business benefits with a shift towards a service perspective. With the Transformate phase, ICT evaluation began to focus more on the intangible aspects of business benefits, including issues such as trust, loyalty, and brand improvement in evaluation frameworks.

Before we explore the possible parallels of these models in the field of ICTD evaluation, let us examine in more detail, the intangible benefits of IT in the business world.

Intangible Benefits of IT in Business

In the business world, especially from the Transformate perspective described above, there are various perceptions of the value that businesses derive from ICT. The value

¹⁰⁶ Laudon, K.C. & Laudon, J.P. 2000. Management information systems: organization and technology in the networked enterprise. 6th ed. Prentice Hall, Upper Saddle River, NJ.

¹⁰⁷ Zuboff, S. 1988. In the age of the smart machine: the future of work and power. Basic Books, New York.

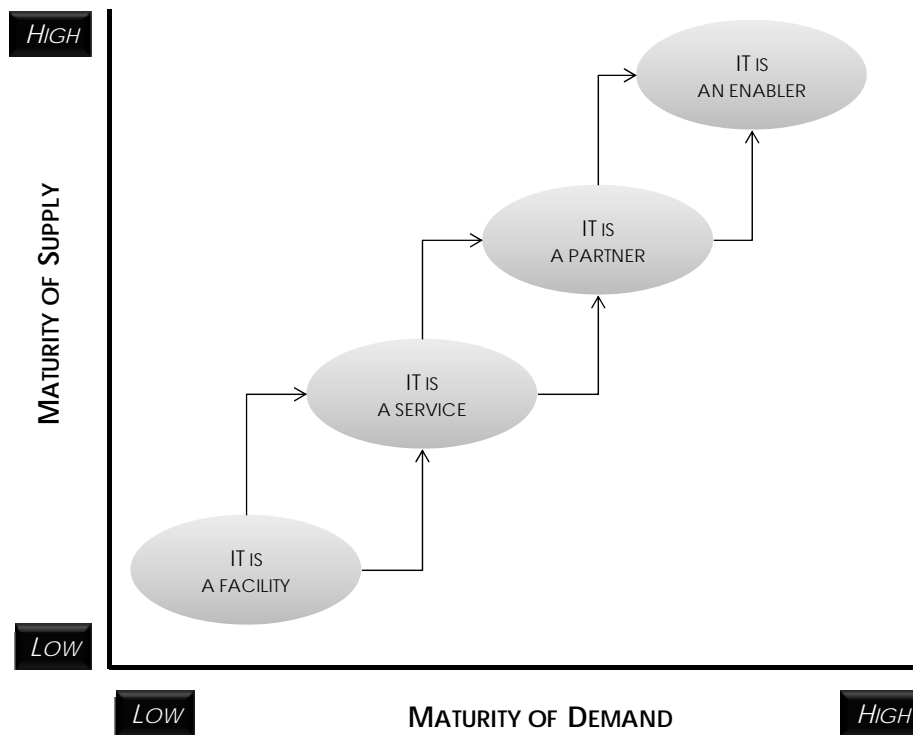
placed in ICT is seen to become higher as its use in the organization progresses from being just a facility to that of an enabler. Wiggers and colleagues suggested the IT Value Perception Model to describe how IT value is perceived:

“The maturity of IT supply deals with the professionalism and quality of the IT function in the organization, with maturity being measured using a quality model such as the Capability Maturity Model (CMM). Maturity of IT demand refers to the self-awareness and self-consciousness of businesses to use and demand an appropriate level of quality from their IT supportive organizations. The IT value perception describes the perception of the executive management of the added value that IT delivers to the company”¹⁰⁸.

Similarly, in the ICTD environment as both maturity of demand and supply increase with increasing investments in infrastructure, our evaluation foci should be tending towards dimensions and metrics related to enablement. Thus, the context of ICT from a development perspective must focus on ICT as an enabler i.e. of socio-economic development. However, this is exactly where the evaluation challenge becomes more complex, as ICT may enable a diverse set of outcomes, which are difficult to link from an attribution or cause-effect perspective.

¹⁰⁸ Wiggers, P., Kok, H. & De Boer-De Wit, M. 2004. IT performance management. Elsevier Butterworth-Heinemann, Oxford.

THE IT VALUE PERCEPTION MODEL¹⁰⁹



The implementation of ICTs, regardless of the setting, can result in a variety of benefits, some tangible, some intangible, and even some unexpected benefits.¹¹⁰ In the business domain, a tangible benefit is one which directly affects the firm's profitability, whereas an intangible benefit is one which can be seen to have a positive effect on the firm's business, but does not necessarily influence the firm's profitability directly.¹¹¹ Hitt and Brynjolfsson¹¹² posit that the

¹⁰⁹ Source: Wiggers et al., 2004: 4

¹¹⁰ Kohli, R., Sherer, S.A. & Baron, A. 2003. Editorial – IT investment payoff in e-business environments: research issues. *Information Systems Frontiers*, 5(3):239-247.

¹¹¹ Remenyi, D., Money, A.H., Sherwood-Smith, M. & Irani, Z. 2000. *The effective measurement and management of IT costs and benefits*. 2nd ed. Elsevier Butterworth-Heinemann, Oxford.

¹¹² Hitt, L.M. & Brynjolfsson, E. 1996. Productivity, business profitability, and consumer surplus: three different measures of information technology value. *MIS Quarterly*, 20(2):121-142.

question of IT value is not a single one, but rather consists of several related but distinct issues such as increased productivity, i.e. is there now more output per given quantity of input? improved business profitability, i.e. has the business been able to use IT to gain competitive advantage and earn higher profits than it would have earned otherwise? Improved value for consumers, i.e. what is the magnitude of benefits that have been passed on to the consumers? In terms of the latter, the issue of profits is not relevant in the ICTD domain. However, improved value for both individuals and communities is of increasing concern, and as suggested already, are the measurement constructs associated with such value is usually intangible.

EXAMPLES OF TANGIBLE AND INTANGIBLE IT BENEFITS IN A BUSINESS CONTEXT¹¹³

IT BENEFITS IN BUSINESS	TANGIBLE (DIRECTLY AFFECTS FIRM'S PROFITABILITY)	INTANGIBLE (HAS NO DIRECT AFFECT ON PROFITABILITY)
QUANTIFIABLE (CAN BE MEASURED)	MAY BE OBJECTIVELY MEASURED; E.G. INCREASE IN REVENUE; REDUCTION IN COSTS	DIFFICULT TO MEASURE OBJECTIVELY E.G. OBTAINING INFORMATION FASTER; IMPROVED CUSTOMER SATISFACTION
UNQUANTIFIABLE (CANNOT BE MEASURED OR MORE DIFFICULT TO MEASURE)	PRECISE IMPACT ON PROFITABILITY CANNOT BE MEASURED, E.G. BETTER INFORMATION; IMPROVED SECURITY	DIFFICULT TO PUT A FINANCIAL VALUE TO THE BENEFIT, E.G. INCREASED CUSTOMER CONFIDENCE; CUSTOMERS OR EMPLOYEES' PERCEPTION OF THE FIRMS PRODUCT.

We can observe that unquantifiable, intangible benefits are those, which are the most difficult to measure. However, this does not imply that these areas of intangible IT benefits need to be excluded from evaluation. Likewise in the ICTD domain we are hard pressed to extend evaluation frame-

¹¹³ Adapted from: Remenyi et al., 2000: 29-30, 152-153

works of quantifiable data that are more easily measured such as teledensity, bandwidth per capita, number of connection points, number of training certificates issued etc. Taylor & Zhang¹¹⁴ hone in on the issue by arguing that “when a technology is regarded as the prime initiator of change in society, measuring the changing technology might seem to be enough” and that “measuring computers, cables, and connections tells us very little about the actual state of society”.

Let us now turn to the evaluation of ICT in development contexts. In doing so we draw some lessons from the business world in order to help move the field from the early euphoria, subsequent instrumental focus on primarily tangible, economic benefits of ICT on society, towards a more “mature” evaluation perspective which incorporates intangible, unquantifiable impacts of ICT in development contexts.

Evaluation of ICT in Development Contexts

Although we have developed a substantive body of knowledge regarding the evaluation of ICTs in business contexts, we have yet to understand how the extant research outputs can be applied in ICTD contexts. Researchers, studying business benefits of ICTs, have over the years adapted theories from other disciplines in their quest to develop models for evaluating ICT benefits. Examples of these

¹¹⁴ Taylor, R. & Zhang, B. 2007. Measuring the impact of ICT: theories of information and development. Telecommunications Policy Research Conference, September 26-28, 2007, Washington, D.C.
http://www.intramis.net/TPRC_files/TPRC%2008%20Taylor-Zhang%20Final.pdf

include Communications Theory,¹¹⁵ Resource-Based Theory,¹¹⁶ and Theory of Reasoned Action.¹¹⁷ However, even the business information systems researchers have been challenged. This is evident in the fragmented body of ICT effectiveness knowledge and lack of consistent trends in the application of the associated theoretical paradigms. Moreover, the tools and techniques to evaluate ICTs in the business environment are not useful in the context of societal development. This therefore exacerbates the problem, as existing theories and models need to be studied carefully and adapted if this is indeed possible, for application in the context of ICTD evaluation. Alternatively new theories for ICT evaluation needs to be developed. This of course has been recognized by the scientific community, and there have been a number of calls to the ICTD research community to respond to the challenge by providing new theories and methods e.g. van Dijk (2006); Heeks (2006); Pather & Uys (2010)¹¹⁸.

A new approach to ICT evaluation in a development context would also be convergent with transformations in the notion of development itself. While early notions of development simply equated it with economic growth and transfer of technology from developed countries to underdeveloped ones, theories of development have long abandoned such simplistic and mechanistic approaches in favour of a more

¹¹⁵ Shannon, C.E. & W. Weaver. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, IL.

¹¹⁶ Penrose, E.T. 1959. The theory of the growth of the firm. John Wiley, New York.

¹¹⁷ Fishbein, M. & I. Ajzen. 1975. Belief, attitude, intention and behaviour: an introduction to theory and research. Addison-Wesley, Reading.

¹¹⁸ Pather, S. & Uys, C. 2010. A Strategy for Evaluating Socio-Economic Outcomes of an ICT4D Programme. Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS-43), CD-ROM, IEEE Computer Society, January 2010.; Heeks, R. 2006. Theorizing ICT4D Research. Information Technologies and International Development, 3(3): 1-4.

holistic view. Such a view includes meeting basic needs in an endogenous process that builds participatory democracy, strengthens self-reliance, promotes structural changes, and fosters empowerment and liberation.¹¹⁹ However, the changes in approaches and theories of development do not seem to have affected the field of ICTD, which seems to be still in the development as modernization paradigm.

Today, thirty years after the ICT productivity paradox was highlighted, the challenges faced in evaluating the impact and productivity of are as relevant in the ICTD context as they were in the business environments back then. This underscores the difficulties associated with measuring ICTs regardless of the area of application and not least of which relates to the issue of impact measurement. The ‘productivity’ resulting from ICTs in respect of facilitating socio economic development of the large numbers of impoverished and underserved communities are still to be properly understood. This is underscored by Heeks,¹²⁰ who in evaluating the initial ICTD era argues that insofar as evaluation is concerned, the work in this field was “held aloft by hype and uncorroborated stories, which fostered a new interest in objective impact evaluation”.

Thus, in the ICTD environment the same questions that confronted businesses for many years still prevail, especially since the digital divide is a phenomenon linked not only to the topic of access to the Internet, but also intrinsically to the one of usage and usage benefit.¹²¹ Even though millions of

¹¹⁹ Servaes, J. 2008. *Communication for Development and Social Change*. Sage, New Delhi.; Melkote, S. 2001. *Communication for Development in the Third World*. Sage, New Delhi.

¹²⁰ Heeks, R. 2008. ICT4D 2.0: The next phase of applying ICT for international development. *Computer*, 41(6): 26-33.

¹²¹ Fuchs, C., & Horak, E. 2008. Africa and the digital divide. *Telematics and Informatics*, 25(2): 99-116.

dollars have been spent by donor and government agencies around the world on ICTs, we still do not have sufficient insight into appropriate methods for evaluating the effectiveness of these technologies on especially socio-economic development. Many ICTD studies and commissioned research by governments tend to focus on quantitative data in respect of penetration and rates of usage and adoption. However, the true value of social development is not easy to conceptualize and hence measure. Parthasarathy & Srinivasan¹²² make a strong case that by relying on easy to measure data for well-defined indicators, econometric techniques may well suffice to measure development, but development also leads to changes that are not economic and not all social changes lend themselves to measurement using well-defined indicators. Qureshi¹²³ is also critical of macroeconomic models used by International Agencies to predict the effects of government policies relating to information technology investments and services on economic growth. She argues that while these models play a pivotal role in decision-making, they often cannot explain why certain IT policies do not have the effects intended, or why certain investments in IT infrastructure do not bring about social and economic change. Taylor & Zhang lend support to this critique when they assert that ICTs do not create the transformations in society by themselves; they are designed and implemented by people in their social, economic, and technological contexts.

¹²² Parthasarathy, B. & Srinivasan, J. 2006. Innovation and its Social Impacts: The Role of Ethnography in the Evaluation and Assessment of ICTD Projects. Global Network for Economics of Learning, Innovation, and Competence Building Systems Conference, 4-7 October, Trivandrum, India.

¹²³ Qureshi, S. 2005. How Does Information Technology Effect Development? Integrating Theory and Practice into a Process Model. Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, USA, August 11th-14th: 500-509.

Tangible Impacts of ICT for Development

There are no dissenting voices in respect of the potential of ICTs to support social development and numerous interventions have been deployed around the world with the goal of bridging the so-called “digital divide”. Governments, various non-government and private for-profit organizations have jointly invested significantly in this effort. Heeks, for example, in examining a “very broad notion” of ICTD estimates that in 2007 US\$840 billion has been invested in all developing and transitional economies, and approximately US\$57 billion in low-income countries (GNI <US\$935). Notwithstanding such astronomical figures of investment, well-articulated evaluation frameworks, spanning both tangible and intangible impacts, are still a major obstacle common across developed and developing countries.

A variety of factors account for this problem, including the sheer complexity of the task of measuring the impact of information on development, the lack of appropriate indicators of impact, and the difficulty of collecting relevant data. Ten years ago we showed that “until relevant methodologies and adequate tools are developed to effectively assess the social impact of the application of ICTs for sustainable development from the user’s perspective, efforts to demonstrate how people are empowered by knowledge will lack credibility”¹²⁴. The evaluation problems were also corroborated by others, such as Batchelor¹²⁵ whom, in reporting in a study of seventeen of the World Bank’s

¹²⁴ Gomez, R., Hunt, P., & Lamoureux, E. 1999. Telecentre Evaluation and Research: a global perspective. In R. Gomez & P. Hunt (Eds.), *Telecentre Evaluation: a global perspective*, IDRC, Ottawa: 13-29.

¹²⁵ Batchelor, S., Evangelista, S., Hearn, S., Peirce, M., Sugden, S. & Webb, M.. 2003. *ICT For Development: Contributing to the Millennium Development Goals – Lessons Learnt from Seventeen infoDev projects*. World Bank, Washington DC.

InfoDev ICTD projects found that, "...analysts and decision makers are still struggling to make sense of the mixed experience of information technologies in developing countries. Very often, such experiences seem to amount to little more than a heterogeneous and unrelated set of anecdotes. However spectacular, successful, moving, or important some of these anecdotes may be, they remain a precarious basis for justifying major policy or investment decisions".

Indeed, there can be no doubt that we have witnessed steady, albeit slow, progress insofar as closing the digital divide from the perspective of broad access to ICT. For example, ITU statistics show that over the last 10 years, the digital divide between the developing and the developed countries has been narrowing in terms of fixed telephone lines, mobile subscribers, and Internet users. In contrast to the slow fixed line growth, phenomenal growth rates in the mobile sector particularly, have been able to reduce the gap that separates the developed from the developing countries from 27 in 1994, to four in 2004. The fixed line gap has been reduced from 11 to four during the same period¹²⁶.

Nonetheless, more infrastructure and more access does not necessarily translate into more meaningful use of ICT, and the evidence of its contribution to social and economic development is still limited. International agencies have explored innovative ways to measure the success of their initiatives and their impact on human development with different degrees of success. Let us briefly discuss some of the most salient experiences in this regard:

¹²⁶ ITU. 2006. World telecommunication/ICT development report 2006 - Measuring ICT for Social and Economic Development, Executive Summary. International Telecommunications Union, Geneva.
http://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-WTDR-2006-SUM-PDF-E.pdf

The World Bank

The World Bank and its affiliates, the International Telecommunications Union (ITU) and the OECD have tended to focus on statistical indicators of economic growth and human development indices. The World Bank's recent studies focus on the deployment of broadband connectivity and mobile telephony, and its contribution to businesses and government applications. (World Bank, 2009 ICTD Report, *Extending Reach and Increasing Impact*).

The World Bank's flagship program on ICTD, InfoDev, has focused on drawing implementation lessons from its many pilot projects around the world, and developed a framework for monitoring and evaluation of pilot projects on ICTD¹²⁷.

International Donors

Even though several international donors (Canadian, Swiss, Danish and Swedish International Development agencies (CIDA, SDC, Danida and Sida, respectively) and international agencies (UNESCO in particular) have followed broadly the same direction of InfoDev and other World Bank program directions described above, agencies such as the Canadian International Development Research Centre (IDRC), on the other hand, have increasingly moved away from a focus on impact to that of outcomes in evaluating development activities¹²⁸. The shift is significant: rather than trying to assess large-scale social impacts, which happen slowly over time and are the result of what many people do,

¹²⁷ Batchelor, S. & Norrish, P. (undated). Framework for the assessment of ICT pilot projects: Beyond Monitoring and Evaluation to Applied Research. InfoDev, World Bank, Washington DC. <http://www.infodev.org/en/Publication.4.html>

¹²⁸ Earl, S., Carden, F. & Smutylo, T. 2001. Outcome mapping: building learning and reflection into development programs. IDRC, Ottawa.

outcome mapping focuses on assessing the direct and observable changes in behaviour that can be directly attributed to the activities that are implemented as part of a development program. The shift from evaluating impact to mapping outcomes comes at a cost: the detailed process to effectively conduct outcome mapping is extremely time consuming, and the resulting findings offer limited explanation of the larger, broader or longer-term benefits social development is expected to produce.

The evaluation of ICTD activities has been caught at the crossroads of these three trends. The statistics and tangible indicators for economic growth, the lessons learned from pilot studies or scattered anecdotes of success, and the detailed documentation of outputs and outcomes do not add up to an understanding of larger social or economic impacts, which are essential to inform better policy decisions and program implementation.

A Way Forward: Assessing Intangibles in ICTD

A fourth trend is emerging in the ICTD evaluation field. This trend is demonstrative of evaluation frameworks, which focus less on direct and tangible economic, social, or political benefits and more on indirect, intangible benefits. These are more difficult to measure and point to deeper and more pervasive effects on individuals and communities.

After 20 years of euphoria, and after 10 years of seeking out tangible, quantifiable, and instrumental impacts of ICT for development, we are still missing the boat: we are still unable to account for the impacts (both positive and negative) of ICT for human development. It may be time to start

building a framework that allows us to evaluate the intangible, non-instrumental, unquantifiable impacts of ICT in development contexts.

Drawing a parallel with the business models discussed earlier, in the field of ICT for development we appear to be still in the early phases of measuring tangible, quantifiable benefits (in Zuboff's model, the Automate phase (measuring technical aspects of ICT) and Informate phase (measuring ICT production and project implementation), but with little inroads into an equivalent of the Transformate phase seen in the business world: one that focuses on intangible, unquantifiable benefits as well as the more easily measured tangible, quantifiable benefits. What would the intangible benefits of ICT for development look like?

PHASES OF ICT EVALUATION IN BUSINESS AND IN DEVELOPMENT

PHASES OF IMPACT EVALUATION	BUSINESS IT (BASED ON ZUBOFF)	ICT AND DEVELOPMENT
<i>EARLY: QUANTIFIABLE, TANGIBLE OUTPUTS</i>	AUTOMATE: MEASURE TECHNICAL ASPECTS OF IT (RATE OF INFORMATION FLOW, ACCURACY, TIMELINESS)	EUPHORIA: QUANTIFIABLE OUTPUTS (NUMBER OF COMPUTERS, NUMBER OF USERS, TELEDENSITY ETC)
<i>INTERMEDIATE: MEASURABLE, MOSTLY TANGIBLE OUTCOMES</i>	INFORMATE: IT PRODUCTION AND PROJECT IMPLEMENTATION	INSTRUMENTAL: ECONOMIC OUTCOMES (INCOME GENERATION, BUSINESS OPPORTUNITIES, ETC)
<i>MATURE: UNQUANTIFIABLE, INTANGIBLE IMPACTS</i>	TRANSFORMATE: SERVICE PERSPECTIVE, INTANGIBLE BENEFITS (TRUST, LOYALTY, BRAND, ETC)	INTANGIBLES: ASPIRATIONAL OUTCOMES (EMPOWERMENT, SELF ESTEEM & SELF WORTH; SOCIAL COHESION & SOCIAL FABRIC)

There are indications of a growing awareness of the need to explore intangible and unquantifiable impacts, even among the international organizations that have promoted the early focus on tangible and measurable outputs and outcomes. For

example, the World Telecommunication and ICT Development report (ITU, 2006) stated that:

Instead, the most important economic impact of the spread and use of ICTs is indirect, by transforming the way individuals, businesses and other parts of the society work, communicate, and interact. The beneficial impact of ICTs on productivity – which can help reduce poverty – is of particular interest as ICT diffusion levels across all countries rise. One way of understanding the difficulty of measuring the impact that ICTs have, is to imagine the impact that electricity has had on the economy and society. As with ICTs, there is no denying that electricity has had important impacts on individuals, businesses and society at large but its measurement is elusive. Part of the difficulty is that both ICTs and electricity are “enabling” or “General Purpose Technologies”, which means their use and their impacts are ubiquitous yet difficult to measure because they are mainly indirect. It is not electricity or ICTs as such that make the (bulk) impact on economy and society but how they are used to transform organization, processes and behaviours.

The foregoing endorses our views regarding the “elusive” and “ubiquitous” nature of ICTD impact measurement, and also highlights the importance of focusing on the “enabling” and “transformational” aspects in our evaluation frameworks. McNamara¹²⁹ extends this point in the following excerpt, in emphasizing the enabling role of ICTs in one of these elusive areas of impact viz. that of social capital:

There is increasing evidence that a dense and complex layer of social institutions, formal and informal groups, and

¹²⁹ McNamara, K.S. 2003. Information and Communication Technologies, Poverty and Development: Learning from Experience. InfoDev, World Bank, Washington DC.

networks of interaction and common interest between the individual citizen and the state is good both for the stability and responsiveness of the political system and for the economy and society as a whole. This social capital enables richer and more diverse views to surface on important societal issues, empowers groups to address common concerns and interests without necessarily relying on government intervention, and can even help the emergence of social consensus by permitting multidirectional debate and sharing of information among those with different perspectives through mechanisms that are not directly tied to the formal political and governmental structure. By facilitating new forms of many-to-many communication, collaboration, and information-sharing, both within a given country and among groups with similar interests and concerns across borders, ICTs can add to the vibrancy of civil society institutions and networks as a check on government, a source of ideas and innovations, and an outlet for the interests, concerns and desires for solidarity on the part of individuals and groups.¹³⁰

A question worthy of introspection is “Why have these calls for attention to intangible impacts, from influential people and organizations, fallen on deaf ears?” Our premise is that this lack of response is simply because evaluating intangible impacts is far more challenging, and hence we have chosen not to go down the path of complexity and murkiness, thus our baskets of tangible and quantifiable indicators remain, as they are, incomplete as they may be. Similar evaluation problems have been faced before e.g. evaluating benefits of water supply and electricity. In one such study highlights the intangible aspect of empowerment. In this study, it was

¹³⁰ (McNamara, 2003)

found that given that the poor often feel marginalized and their lives can be characterized by a sense of powerlessness and instability, even the perception of benefits could assist in empowerment, which is likely to lead to proactive initiatives by the poor themselves (Songco, 2002¹³¹). Another example of an intangible impact, referred to by McNamara above is that of cohesiveness. In the words of Reimer, only cohesive people can “respond collectively to achieve valued outcomes and to deal with the economic, social, political, or environmental stresses (positive or negative) that affect them” (Reimer, 2002¹³²). This therefore suggests that the tangible impacts that we have been so focused on, are in fact facilitated by and preceded by the intangible impacts on which we have been focused. Thus, we have expended our energies on evaluating impacts which are too far down the impact timeline, and which has contributed to the fuzziness, which belies the linking of ICTs with development.

We suggest therefore that examples of intangible impacts which need to be evaluated could include empowerment, self-esteem and sense of self-worth, at the individual level, and social cohesion and strengthening of social fabric, at the collective level. Social cohesion was defined by Ritzen¹³³ and colleagues as: “a state of affairs in which a group of people demonstrate an aptitude for collaboration that produces a climate for change”. According to these authors, the role of social cohesion is to strengthen the trust inside and between

¹³¹ Songco, J.A. 2002. Do Rural Infrastructure Investments Benefit the Poor? Evaluating Linkages: A Global View, A Focus on Vietnam. World Bank, Washington DC.

¹³² Reimer, B. 2002. Understanding and Measuring Social Capital and Social Cohesion. Montréal: Canadian Rural Restructuring Foundation, Concordia University. Unpublished.

¹³³ Ritzen, J., Easterly, W. & Woolcock, M. 2000. On “Good” Politicians and “Bad” Policies: Social Cohesion, Institutions, and Growth. World Bank Policy Research Working Paper 2448, World Bank, Washington DC.

different social groups in order to accelerate socio-economic growth with “a reasonable distribution over those groups in the population”. In the same way, Maxwell¹³⁴ had identified social cohesion with the processes of building shared values and communities of interpretation, reducing disparities in wealth and income, enabling people to have a sense that they are engaged in a common enterprise, facing shared challenges, and that they are members of the same community. These are all impacts that have been anecdotally reported in the ICTD literature, but with little traction in the face of the dominant attention paid to tangible and quantifiable, economic impacts.

Conclusion

We require new measurement strategies, and a shift in focus from measuring the tangible symptoms or outputs to measuring intangible outcomes i.e. the more strategic development goals such as improved self-reliance and social cohesion. These are important building blocks of socio-economic development, and thus our ICT evaluation toolkits need to account for such indicators. Currently agencies and governments find it difficult to rationalize their expenditure and evidence. At most, much of the evaluation done to date is based on quantitative output. Some agencies, though not all, are able to provide data such as the number of people trained or rate of usage of centres. However, this in itself does not inform us if any meaningful change has been brought about in people’s lives. An important question that should therefore be placed on the ICTD research agenda is:

¹³⁴ Maxwell, J., 1996. Social Dimensions of Economic Growth. Eris John Hanson Memorial Lecture Series, Vol. VIII, University of Alberta, Canada.

How do we extract credible evidence of the socio-economic change in people's lives, especially when this change is unquantifiable and intangible? The value of ICTD evaluation lies in producing data that substantiates socio-economic outcomes and impact. The research and practitioner community has to work hard to ensure that as the next era, viz. ICTD 2.0 (see Heeks, 2008) dawns upon us we should no more held aloft by hype and uncorroborated stories.

Let us take the case of a small semi-rural fishing village in the southernmost tip of Africa. None of the anglers in this village have actually learned how to use the public access computer at the local library, a typical community based ICTD project. However, evidence from a case study (Pather & Mitrovic, 2008¹³⁵) indicates that the public access computers in the library provided a means of uniting the anglers around a common cause related to fishing licences. This in turn improved their basic livelihoods in indirect ways. If the evaluation of the centre was based on tangible measures of usage statistics, such an impact would not have been identified. In this case, a closer look at issues of empowerment and cohesion revealed evidence that ICTs have indeed transformed the life of the community in ways that are far more important than what the number of users of computers would have indicated.

This example is a typical case of an anecdote that does not build up to evidence of impact for lack of a framework that goes beyond tangible and quantifiable indicators of success of ICT for development projects. A collection of numerous anecdotes such as this does not, in itself, constitute evidence of impact. Instead, a structured approach to evaluate impact

¹³⁵ Pather, S. & Mitrovic, Z. 2008. ICTs for participatory local development in South Africa: Exploring a systemic approach. IT for Change, Bangalore, India.

of ICT, which includes attention to intangible and unquantifiable aspects (e.g. self-worth and the strengthening of social fabric) that are facilitated indirectly through use, or even the presence of ICTs, will provide a more complete and holistic perspective of ICT impact.

Finally, we discussed how the euphoria and disillusionment with ICT for development might be a consequence of two issues. Firstly, it stems from its roots in a modernization paradigm, which equates development with economic growth. Secondly, it has to do with an evaluation approach that focuses on measuring quantifiable, tangible outputs and outcomes of the ICTD interventions. Drawing from the experience of ICT evaluation in the business environment, we suggest that it is not only possible but also desirable to mature into an evaluation approach, which includes attention to intangible and unquantifiable impacts of ICT. Thus, a focus on empowerment and social fabric instead of economic growth would also be convergent with current theories and approaches to development, which have long abandoned the modernization paradigm. To this end, all stakeholders in the ICTD environment, including researchers, practitioners, and funders need to undergo a process of conscientization. In so doing, we should aim to detach ourselves from the fixation with narrow tangible and quantitative impact metrics. A more detailed exploration of the theoretical and methodological implications of this shift in the ICT for development field is warranted.

CHAPTER NINE

Evaluating Gender Equality: Universalia Experiences and Perspectives

*Katherine Garven, Elisabetta Micaro, Katrina Rojas,
and Anette Wenderoth*

Universalia is not a firm that solely specializes in gender. However, over the past years, there has been a significant increase in the proportion of our evaluation work related to gender issues. Many of the evaluations we are asked to carry out either contain a gender component, as a crosscutting issue, or focus on gender equality as the primary theme of the evaluation. This increase appears to be, at least partly, due to the evolving international development context and its changing evaluation needs. Our approach to gender equality treats the topic through an evaluation lens rather than through the perspective of gender experts and this has proven informative and helpful to many clients.

Among the international development community, gender equality is increasingly considered as crucial, as both a means to and an outcome of development. UNESCO, for example, states, “our work is guided by the premise that the promotion of gender equality is both a means for develop-

ment and an end for the enjoyment of human rights by women and men".¹³⁶ Gender equality plays a significant role within the framework of the Millennium Development Goals (MDGs). Gender equality is explicitly mentioned in Goal #3: Promote gender equality and empower women as well as implicitly in Goal #5: Improve maternal health.¹³⁷ In fact, many activists believe that gender equality is indeed crucial in achieving all eight MDGs. The UNDP has stated "gender equality and women's empowerment are human rights that lie at the heart of development and the achievement of the Millennium Development Goals".¹³⁸

As gender takes a more central role in international development, organizations have begun to include gender issues into their program and organizational frameworks. In fact, many players have established policies to mainstream gender into all of their program and organizational activities. The adjustment of evaluation norms and standards is also taking place. For example, the United Nations Evaluation Group (UNEG) is currently preparing additional guidance for UN agencies on how to address gender in evaluation.¹³⁹ Consequently, organizations are becoming more and more interested in understanding how well they are addressing gender equality. Universalia is, therefore, increasingly asked to include gender equality as a crosscutting issue into its evaluation work.

¹³⁶ UNESCO Gender Equity and Development
http://portal.unesco.org/shs/en/ev.php-URL_ID=1375&URL_DO=DO_TOPIC&URL_SECTION=201.html

¹³⁷ United Nations Millennium Development Goals
<http://www.un.org/millenniumgoals/gender.shtml>

¹³⁸ UNDP Gender Equality and Women's Empowerment
<http://www.undp.org/women>

¹³⁹ UNEG "Guidance on Integrating Human Rights and Gender Equality Perspectives in Evaluations in the UN System." 2009.
www.unevaluation.org/documentdownload?doc_id=395&file_id=584

This chapter aims to serve as a platform for reflection upon our experiences to date in evaluating gender issues and for considering how to best move forward in effectively evaluating gender equality in the future. The chapter is based on the experiences of various Universalialia consultants in evaluation and project implementation work. It is divided into six sections. The second section following this introduction briefly outlines some gender specific terms and definitions; the third discusses Universalialia's gender-related experience and the firm's evolution in evaluating gender; the fourth section outlines some challenges that we have faced in evaluating gender; the fifth presents some ways in which we have attempted to mitigate these challenges; and the final section concludes with some forward looking reflections.

Key Definitions

Gender is a highly complex social construct. As a result, the term gender is still a comparatively under defined concept. At the multilateral level, states have generally avoided defining gender due to the lack of a consensus about the precise content of the term. Often the term is either undefined or the respective authors merely state that the term is to be interpreted or understood "in its ordinary, generally accepted usage".¹⁴⁰

The United Nations emphasizes that gender is a social construct, in comparison to the term sex, which is defined as "the biological differences between men and women, which

¹⁴⁰ Oosterveld, Valerie "The Definition of "Gender" in the Rome Statute of the International Criminal Court: A Step Forward or Back for International Criminal Justice?" Harvard Human Rights Journal Vol. 18, Spring 2005
<http://www.law.harvard.edu/students/orgs/hrj/iss18/oosterveld.shtml#Heading99>

are universal, obvious, and generally permanent.”¹⁴¹ The most commonly-used definition of the term gender, which is the one used throughout this chapter, comes from the Office of the Special Advisor on Gender Issues and the Advancement of Women in the Department of Economic and Social Affairs (OSAGI) at the United Nations. It defines gender as “the social attributes and opportunities associated with being male and female and the relationships between women and men and girls and boys, as well as the relations between women and those between men. These attributes, opportunities, and relationships are socially constructed and are learned through socialization processes. They are context/ time-specific and changeable. Gender determines what is expected, allowed, and valued in a women or a man in a given context. In most societies, there are differences and inequalities between women and men in responsibilities assigned, activities undertaken, access to and control over resources, as well as decision-making opportunities. Gender is part of the broader socio-cultural context. Other important criteria for socio-cultural analysis include class, race, poverty level, ethnic group, and age.”¹⁴²

There is wide agreement that the social construct of gender has lead to (or has at least significantly contributed to) social, political, economic, and cultural inequalities between men and women. In recent years, the pursuit of gender equality and gender equity through strategies such as gender mainstreaming has become an international priority. The most commonly used definition for gender equality comes from the OSAGI, who defines gender equality as “the equal

¹⁴¹ UNDP Gender Mainstreaming Definitions

¹⁴² OSAGI (Office of the Special Advisor on Gender Issues and the Advancement of Women Department of Economic and Social Affairs (UN)
<http://www.un.org/womenwatch/osagi/conceptsanddefinitions.htm>

rights, responsibilities and opportunities of women and men and girls and boys. Equality does not mean that women and men will become the same but that women's and men's rights, responsibilities and opportunities will not depend on whether they are born male or female. Gender equality implies that the interests, needs, and priorities of both women and men are taken into consideration, recognizing the diversity of different groups of women and men. Gender equality is not a women's issue, but should concern and fully engage men as well as women. Equality between women and men is seen both as a human rights issue and as a precondition for, and indicator of, sustainable people-centered development".¹⁴³

The term gender equity "means fairness of treatment for women and men, according to their respective needs. This may include equal treatment or treatment that is different but which is considered equivalent in terms of rights, benefits, obligations, and opportunities."¹⁴⁴ To ensure fairness, measures must often be available to compensate for historical and social disadvantages that prevent women and men from otherwise operating on a level playing field. For example, a gender equity program could deliver special care to pregnant women. Even though the same program cannot be delivered to men (because men are incapable of being pregnant), the program is equitable in that both men and

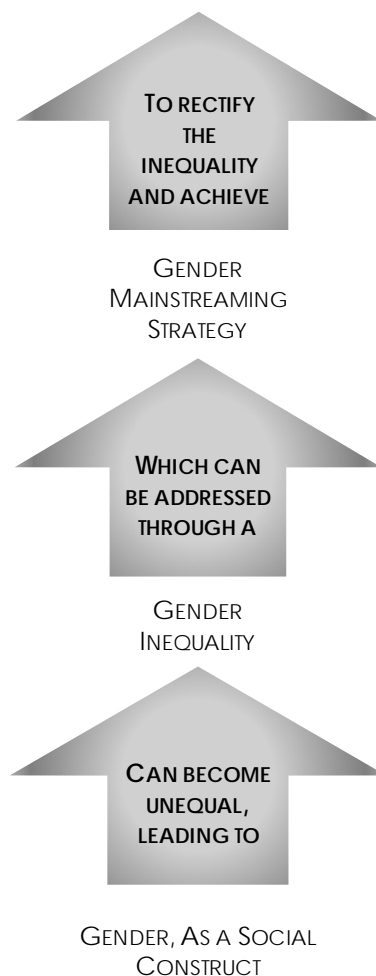
¹⁴³ OSAGI (Office of the Special Advisor on Gender Issues and the Advancement of Women Department of Economic and Social Affairs (UN) <http://www.un.org/womenwatch/osagi/conceptsanddefinitions.htm>

¹⁴⁴ "ABC of Women Worker's Rights and Gender Equality". International Labour Organisation. Geneva. 2000. P. 48

women will enjoy good health. Equity is expected to lead to equality.¹⁴⁵

RELATIONSHIP BETWEEN THE PREVIOUSLY DEFINED GENDER TERMS

GENDER EQUITY AND GENDER EQUALITY



Gender mainstreaming is a common strategy, which can be used to make progress towards gender equality. At the World Conference on Women in Beijing in 1995, the concept

¹⁴⁵ McLaren, Joan "Evaluating Programs for Women: A Gender-Specific Framework" Prairie Women's Health Centre of Excellence Research Policy Community 2000 <http://www.pwhce.ca/pdf/evaluatingPrograms.pdf>

of gender mainstreaming was internationally defined in the Beijing Platform for Action as “Mainstreaming a gender perspective is the process of assessing the implications for women and men of any planned action, including legislation, policies or programs, in all areas and at all levels. It is a strategy for making women's as well as men's concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of policies and programs in all political, economic and societal spheres so that women and men benefit equally and inequality is not perpetuated. The ultimate goal is to achieve gender equality”.¹⁴⁶

The concept of gender mainstreaming is at the core of the idea of viewing gender as a crosscutting issue. The idea of integrating gender concerns into the program and organizational frameworks of development organizations has led to the inclusion of gender as a crosscutting issue into program and organizational evaluations. It should be noted, however, that gender mainstreaming is only one of many possible ways of working towards gender equity and equality and is often highly criticized for its vaguely defined operationalization and its tendency to weaken accountability by making gender everyone's responsibility in theory, which in practice often means that it ends up being no one's responsibility.

Universalialia's Work and its Evolution in Evaluating Gender

Although Universalialia does not consider itself a gender-specialized firm, we have been able to gain considerable and interesting gender-related experiences over the past 30 years.

¹⁴⁶ The United Nations Report of the Economic and Social Council for 1997

During that time, Universalialia itself has evolved into a more gender-sensitive firm. Currently, there is a balance between the sexes among the current senior leadership and general staff, with both perspectives reflecting insights on gender equality.

Universalialia's work has typically focused around project and program evaluations, organizational assessments, and the implementation and management of projects. Several of our clients who have asked us to look at gender issues are agencies that themselves are taking a leadership role in pushing the gender equality agenda. The Canadian International Development Agency (CIDA) is an agency, which is well known for prioritizing gender as a development issue and is widely considered as a leader in this regard among donor agencies. Universalialia has had a successful working relationship with CIDA since the firm's establishment. Another of our clients who is pushing the gender agenda is the United Nations Development Fund for Women (UNIFEM), with whom we have more recently been sharing a very strong working relationship. Other clients who are making efforts to mainstream gender include groups like the United Nations, the African and Caribbean Development Banks, and the World Bank, as well as various international and Canadian NGOs.

Initially, most of our evaluation work at Universalialia focused on gender as a crosscutting issue. However beginning in 1991, we started to conduct more assignments with a stronger focus on gender equality. In 1991, Universalialia was contracted to review CIDA's gender policy and develop a policy framework. Initial scepticism within the agency eventually transformed into widespread interest and enthusiasm for a gender equality policy framework. However, despite the strong internal support among CIDA staff,

there were significant challenges in internalizing the framework to ensure gender equality within the agency itself. Although CIDA's internal operations are today much more gender-sensitive, this project was the first time that we closely witnessed how the mere existence of a gender equality policy does not automatically translate into action or actual changes within an organization. Another key moment which encouraged us to reflect about gender was when we jointly published a book with the International Development Research Centre (IDRC) on Organizational Assessment. While the book was being circulated for feedback, *Universal* was criticized by gender groups for not including a feminist perspective of Organizational Assessment. This provided consultants with an opportunity to reflect on how they incorporate gender issues into their work and how to incorporate such issues. Most recently, our work with UNIFEM has provided many opportunities to reflect upon gender equality issues and their role in international development. Different from most of our other clients, UNIFEM's work is solely dedicated to the promotion of gender equality and women's human rights.

Challenges in Evaluating Gender Equality

Over the past 30 years, we have faced many challenges in evaluating gender. Some of these appear to be specific to gender-related issues, while others can be found in other aspects of our work as well. The following section summarizes some of the key challenges many of our consultants have encountered repeatedly in relation to gender issues.

Challenge 1: The conceptualization of gender equality

The first challenge that we frequently face has to do with the conceptualization of gender. As mentioned previously, gender still is an under defined concept that often causes confusion in the use of its terminology. It is often used in the context of capacity building, which is an equally under-defined term. The use of clear definitions is not only crucial to the evaluation itself, but also to the effective communication of the evaluation findings. Ambiguity and miscommunication can arise if not all parties have the same understanding of relevant concepts. Establishing a clear understanding of gender terms and concepts, which all participants accept, is crucial to a successful evaluation.

Another conceptual challenge lies in the multiple and ever-evolving approaches used to place gender into the development discourse. The place of gender equality in international development has evolved significantly over the past century and continues to evolve. Gender equality was first integrated into the development discourse in the early 1970s as part of the Women in Development (WID) approach. This approach focused on integrating women into ongoing development initiatives but viewed them as passive rather than active recipients of development. In response to the WID approach, a new approach was developed in the mid 1970s, known as the Women and Development (WAD) approach, which focuses on the relationship between women and the development process. The most common approach used today was developed in the 1980s and is known as the Gender and Development (GAD) approach. The GAD views women as agents of change rather than passive recipients of development assistance and looks at

“the totality of social organization, economic, and political life in order to understand the shaping of particular aspects of society”.¹⁴⁷

What makes these approaches confusing is that nowadays many organizations – at least in theory – follow a GAD approach; however, many actually practice aspects from all three approaches. This can cause confusion in our understanding of how to approach gender as a development issue. The under defined terminology and the ambiguous use of gender in development approaches add layers of complexity to the conceptualization of gender.

Challenge 2: The nature of gender equality

Gender differs from some other evaluation topics due to its highly sensitive nature and its position as part of complex social and cultural processes. Gender is a sensitive theme due to its direct connection to people’s identity. When gender is introduced as a theme for discussion, we have noticed that oftentimes people will become defensive and evasive, assuming that by mentioning the issue, we are implying that something is wrong – and not only ‘wrong in general’, but wrong with them personally, their culture, norms, and values. As one of our evaluation colleagues once mentioned, “It (gender) is a sensitive issue to talk about and discuss, so sometimes people don't want to speak about it or don't reply.” Due to current gender inequality, it is also difficult to access information, which presents views from the suppressed gender. Additionally, at times, we have met people who believe that gender equality is a “Western value” which does not pertain to them. Working with such a

¹⁴⁷ Ibid

socially sensitive topic can be challenging and can lead to difficulties in accessing information.

The other challenge is that gender issues, which form part of very complex social and cultural processes, can be difficult to understand. Identifying the sources of gender inequality can be extremely challenging due to the complexity and subtlety of current and historical relationships between men and women. The fact that gender relations are uniquely constructed in different cultural settings adds another dimension of complication. We have found that it can be difficult for people from outside of a culture to understand the gender relations of that culture. At the same time it can be argued that sometimes the view of an outsider can bring clarity to a situation.

Challenge 3: Gender as a crosscutting issue

The act of mainstreaming gender and including it as a crosscutting issue can often lead to it being sidelined by other themes considered integral to the project. It is increasingly common for organizations to be obliged to mainstream gender into all aspects of their work. However, it is not necessarily always felt by project managers to have particular relevancy to their work. Even if it is considered relevant, its position as a crosscutting issue often results in it being viewed as a secondary rather than a primary concern. Gender can therefore become an add-on issue as opposed to being fully integrated into the evaluation approach.

As an add-on issue, gender components of a project frequently lack clarity regarding content, objectives, and their implications to the overall project. People often view gender as simply a requirement, which must be met without appropriate understanding of the rationale for its inclusion

and possible methods for its integration. Project staff often does not receive the necessary gender training to effectively implement and manage gender issues. This “add-on” approach to the integration of gender issues frequently leads to expected project/ program results being poorly defined (if at all) and therefore difficult to evaluate.

Evaluations of projects or programs that include gender as a crosscutting issue are usually expected to answer a broad variety of different questions, only one or two of which relate to gender. In some cases, the terms of reference have consisted of over forty questions with only one or two devoted to gender. This limits the evaluators’ ability to perform in-depth analysis of progress or actual gender-equality results. As a crosscutting issue we face the risk of merely ticking boxes as in ‘yes they did what they said they would do’ without further exploration of whether what they did was actually relevant and to what extent it was effective.

Also, due to the side-lined approach to gender as a crosscutting issue, consultants can be put in uncomfortable positions where crosscutting gender issues must be addressed in order to ensure the overall quality of their work even if their clients would rather dismiss the issues or are resistant to address them. We have been in situations where our clients have not asked us to address gender issues but while performing field research, we discovered that serious issues of gender inequality were pervasive. In one case, we discovered in an Organizational Assessment that sexual harassment was a key issue within the organization. However, gender was not an issue that the organization was interested in having assessed. This put the consultants in a difficult position where they felt ethically obliged to mention the gender inequality issues within the organization while at the same time knowing that this would displease their clients.

Additionally, evaluators can become frustrated and eventually lose some interest in evaluating gender issues if they are consistently presented as “add-on” issues. Evaluators like to feel that their findings and recommendations will be taken into consideration for future projects and that their work is therefore meaningful. When clients include gender into their work purely as a requirement with little interest in advancing the issue, evaluators can become frustrated or lose interest in evaluating an issue that will be most likely ignored and taken less seriously than other issues.

Challenge 4: Measuring and attributing results

Perhaps the most challenging aspect of evaluating gender equality is the difficulty in measuring and attributing results. Gender equality is a broad and long-term goal whose successful attainment cannot be measured within the scope of a one, two, or three-year project or program. What then can development initiatives of usually limited duration hope to achieve, and what can organizations working on gender equality issues be held accountable for?

Another challenge in this context is that gender relations are highly complex social phenomena making it difficult to attribute progress towards gender equality to any one project or initiative. So many different factors affect gender relations that it is very challenging to identify cause and effect.

This also links to the broader question of how to track and demonstrate progress resulting from (“soft”) advocacy and policy development work. Many gender specific projects focus on advocacy or policy related interventions. However, it is very difficult to attribute social change to specific policies or advocacy work. To use a gender policy as an

example, most people would agree that the approval of a gender policy is an important step in moving towards gender equality. However, one could also argue that the existence of a policy does not yet constitute actual change in the lives of women and men. It is very difficult to assess how the development of a gender policy has led to changes in the relationship between women and men. In addition, most organizations do not have the necessary resources to do long term tracking of the impact of their gender policy or advocacy work to prove that their activities have led to an increase in gender equality.

Challenge 5: Lack of specific methodological or applicable frameworks for evaluating gender

Specific methodologies and frameworks for evaluating gender are not yet well established or well known, but there are some useful principles. The most common long-standing convention is to use gender-disaggregated data, which at least gets the evaluator into the realm of serious enquiry about observed differences of gender-disaggregated groups. However, evaluating gender equality is not yet at the stage where simple conceptual frameworks can keep evaluators on track. Therefore, consultants are left without formal guidance to evaluate the complexities and subtleties of gender issues. Evaluators are forced to rely on general evaluation tools such as the Logical Framework and other results-based frameworks. Although these frameworks are often used to plan and then measure results, it is more difficult to use in situations where gender equality results are difficult to attribute to particular projects or programs.

An example of the challenges associated with tracking results can be seen in a global evaluation that we recently conducted on the elimination of violence against women. During the evaluation, we found many anecdotal results achieved by individual projects. Many of these were small sized projects, implemented in a specific community or targeting a specific group of people (e.g. women living with HIV). However, we had been asked to evaluate the whole program in terms of the overall results achieved to end violence against women. In reporting on the results achieved by the program globally, we found it very difficult to assign the correct weighted importance in the evaluation report to these anecdotal and local changes. It was challenging to attribute project activities to progress towards gender equality at the global level.

Some gender-specific frameworks do exist, such as the Harvard Analytical Framework, the Moser Gender Analysis Framework, and the Social Relations Framework, among others. However, they are intended for research and project planning. It is difficult to adapt these frameworks to the field of Evaluation due to their inability to address pertinent evaluation needs. We have found that one of the common problems with frameworks in general is that even if a framework is relevant for a particular project, it is still very time and energy consuming to adapt the framework to the particular project circumstances.

The good news is that the lack of simplistic tools forces the evaluator to analyze and think about what is going on rather than merely completing a checklist. Thorough and well-conceived analysis is often the best contribution of an evaluator in helping people understand the subtle complexities of gender dynamics.

Challenge 6: Including gender in the evaluation terms of reference

Through our experiences in evaluating gender, we have faced a variety of challenges in addressing gender as part of the Evaluation Terms of Reference. When gender is a crosscutting issue, we have noticed that sometimes the issue appears to be added in as an after-thought. Additionally, we have found that gender-specific questions in the Terms of Reference are at times not constructed in a way that will be useful in providing information that will be helpful in addressing gender inequality or that will provide a holistic understanding of the situation. The Terms of Reference are also often insufficiently detailed or clear to outline clients' expectations.

We acknowledge that it is difficult for evaluation managers to address issues of gender equality, as they are obliged to design Terms of Reference that cover all aspects of a complex project, with gender being only one of many accountability obligations. It is particularly difficult for the evaluation manager to create meaningful questions if the project or program itself has not defined its gender approach, strategy, and expected results. These problems typically originate at the project planning stage where gender is still often marginalized.

Vague or broadly defined Terms of Reference are particularly common in cases where an evaluation is carried out on behalf of a donor agency that is providing funding to a project or program implemented by another organization. A donor agency recently contracted Universalia to conduct an Organizational Assessment of an NGO whose aim is to strengthen civil society. Among the 40 key evaluation questions, the donor agency wished to understand the extent to which the NGO had contributed to women being “better-positioned and equipped to champion governance issues”. Neither the evaluation terms of reference nor the funding agreement between the NGO and donor agency defined the concepts of “better-positioned” and “equipped to champion governance issues”. We were left with many questions such as: Were particular groups of women expected to benefit or was it meant to benefit women in general? Did the donor agency expect to see specific governance issues championed? In addition, governance was only one of the organization’s programming areas. Thus, this one area of focus did not seem to encompass the spectrum of the organization’s work. We have come across similarly broad evaluation questions in other projects.

An interesting observation that we have noticed is that even if you are working with an organization that has a strong focus on gender issues; it cannot automatically be assumed that the group uses a gender-sensitive approach when creating the Evaluation Terms of Reference. We have observed that in some gender-specific projects, project rationale is often assumed rather than proven. This is another example where gender equality may not be adequately integrated into the Evaluation Terms of Reference.

Through discussions with our colleagues, we have learned that other evaluators have noticed a lack of a gender-sensitive approach in projects designed to increase gender equality. One of our colleagues explains her experiences as follows:

While the majority of microfinance clients worldwide are women, as are 70% of the world's population living in poverty, many donors have substantially reduced their focus on gender. It seems that given the high numbers of microfinance clients being served are women, that any other gender measurements or specific efforts to evaluate the gendered impact of microfinance on clients, such as loan utilization, are almost completely dismissed. There is an overriding assumption that because women are being served that goals of equality are being met. However, possible negative impacts, such as loan passing wherein the woman is pressured to take microfinance loans in her name - even possibly taking from multiple microfinance institutions and thus becoming over-indebted- and passing the money onto male family members because men are excluded from eligibility, are not even investigated in programs.

Ways in Which We Have Attempted to Mitigate These Challenges

Challenge 1: The conceptualization of gender equality

In order to address the challenges associated with the conceptualization of gender equality, we have begun to make a point of acknowledging that gender is an under

defined concept with ambiguous terminology, which can lead to multiple interpretations. Through this acknowledgement, we try as often as possible to meet with the client before we begin the evaluation to clarify terminology and expectations in order to avoid miscommunication.

In order to have a good understanding of the evolution of approaches used to integrate gender into the development discourse and remain up-to-date with the evolution of these approaches, we make an effort to stay informed using current academic and popular research and through the establishment and maintenance of strong relationships with gender experts. As most of us are not gender specialists by training, we have had a lot to learn, and continue to learn in and through each new assignment.

Challenge 2: The nature of gender equality

Gender relations are complex social phenomena that vary across cultural settings. While it might be easiest to avoid this complexity by merely using quantitative measures to evaluate gender, such as counting how many men and women are included in and benefit from a program, qualitative measures provide us with a better understanding of factors, which construct and influence gender relations.

Although Universalialia, like many other evaluation firms, often uses quantitative measures to evaluate gender issues, we have been recently striving to increase our use of qualitative measures. To the extent possible, we attempt to use ethnographic approaches, which look at the relationships and attitudes between men and women and the social and cultural context, which affects them. We believe that it is important to not only note gender disparities, but to understand the underlying causes for inequality – and to assess

the relevance and effectiveness of programs or projects within their particular context.

An example of a situation where we attempted to use an ethnographic approach was during one of our projects in the Balkans. We noticed that there was an equal number of female and male teachers in the country but that they were distributed unevenly. There were more female teachers in the cities and more male teachers in the countryside. Although there was no obvious evidence of gender disparity or inequality, this phenomenon interested us. Rather than only reporting on quantitative indicators, we asked the qualitative question: "Why is this happening?" Through investigation, we found that teaching was one of the best-paid and most prestigious jobs in the countryside. Therefore, the majority of the teaching posts were given to men. In the cities, however, there was a greater number of better paid and more prestigious jobs than teaching. Therefore, the men assumed the better positions while the women were given the teaching positions. Although the number of male and female teachers within the country was relatively equal, qualitative analysis indicated that there were some issues of gender inequality.

The sensitive nature of gender makes evaluating gender issues similar to evaluating minority issues. Even when women are given the chance to express their views about a project or program, they often do not feel empowered or comfortable to do so. For instance, we have attended meetings where fifty percent of the people present were women. However, the men did all of the talking. This can make it difficult to access necessary information. When information is accessible, we often must look at it critically because it often does not represent multiple perspectives,

particularly those suffering from gender inequality. Data and written documents are particularly notorious for not presenting all viewpoints. We have found that when working with the issue of gender, it is necessary to meet with the participants or beneficiaries of a project in order to observe local situations and hear the viewpoints of those who are marginalized and/ or oppressed.

Challenge 3: Gender as a crosscutting issue

When gender equality is a crosscutting issue, it is convenient for clients and evaluators alike to minimize its role. This is due primarily to the complexity associated with addressing gender equality and its position as a secondary issue in a crosscutting context. Although time and resources are scarce, it is important to pay proper attention to gender issues if gender is to be successfully mainstreamed.

In order to improve the way our clients address gender as a crosscutting issue, we try to help them by illustrating any absence of gender-sensitive activities, approaches, or progress/ results indicators, and gaps in project/ program results frameworks, and other project documents such as results statements and/or gender strategies. In addition, even if based on limited evidence, and pointed out only in oral debriefs, we feel the need to highlight any areas of incongruence that we observe between program or project activities and results and organizational practices. The very least we can do is identify when there may be organizational dynamics or issues that may limit the organization's potential to sustain gender equality approaches in its programming. Hopefully, through this, we can help organizations to better identify areas of improvement in their approach to mainstreaming gender.

Part of the challenge in evaluating gender as a crosscutting issue is due to the vagueness and uncertainty often faced by evaluators when evaluating crosscutting issues. Gender as a crosscutting issue could be evaluated much more effectively if we had a better understanding of how to evaluate crosscutting issues in general. What is the evaluation of crosscutting issues intended to accomplish? How can we ensure that evaluation findings of crosscutting issues are taken seriously and applied to future activities? How can we maximize the benefits of evaluating crosscutting issues?

Challenge 4: Measuring and attributing results

Measuring and attributing gender results to program and project work is one of the biggest challenges associated with evaluating gender equality. Gender issues form part of intricate social relations that can be affected by numerous factors. Therefore, gender equality initiatives form part of a complex social change process, where measuring and attributing results is very difficult.

Some of Universalialia's most relevant work in the domain of gender equality has been through its support of organizations in tracking gender equality results using a Results-Based-Framework (RBF). Although Universalialia does not specialize in gender, we do have a set of practical skills that can be useful in tracking gender equality results. Although the RBF is difficult to apply to complex social change processes, it can be modified in order to track and demonstrate gender results. To do so, we typically focus on long-term change process and assess the organization's change theory. Additionally, we often use results chain logic and Outcome Mapping concepts for tracking progress rather than achievements.

We try to make recommendations on how the organization could track gender equality related progress more effectively and in a meaningful way. By meaningful we mean not only serving accountability purposes but also actually providing relevant and useful information for the organization and its partners. For example, in the recent past we have repeatedly suggested to clients to try to focus on tracking ‘movement’ rather than results achievement. This links to the idea of Outcome Mapping, which reconstructs change theory and intended changes to provide evidence that the project is moving in the right direction.¹⁴⁸

For example, we have recently suggested to some clients to consider introducing the concept of so called progress markers rather than conventional indicators to reflect whether a project or program is making a positive contribution towards the ultimate goal of gender equality. We believe that the key issue to assess is whether or not progress is being made in the right direction. To answer this question, we are convinced that it is necessary to track results at the levels of access and participation as well as output and outcomes levels.¹⁴⁹

In several cases, we have worked with the client to define gender equality related results ex-post or formulate a previously implicit ‘theory of change’ or ‘intervention logic.’ We attempt to understand the implementing organization’s key assumptions underlying its strategy. We often ask questions such as Where do they see the relevance of gender

¹⁴⁸ For more information regarding Outcome Mapping, visit the IDRC website. http://www.idrc.ca/en/ev-26586-201-1-DO_TOPIC.html

¹⁴⁹ See Earl et al. (2001) and the recommendation section of this report for further elaboration

policies or frameworks in view of broader gender equality related changes? What role and relevance do they attribute to women's political participation? How do they see their particular (usually short or midterm) intervention contributing to or triggering broader gender equality relevant changes?

We have noticed that it is common for donor agencies to expect to see concrete gender equality results from the projects that they fund. However, in several instances there are no concrete results because the project was too small or the organization was only in the process of developing its own internal capacity to address gender equality issues. In these cases, we try to assess the level of gender sensitivity in the project design.

This was, for example, the case in one of Universalia's projects where the consulting team was asked to assess the renovation of schools. Due to the small-scale nature of the project, we were unable to identify any concrete gender results. Therefore, we tried to assess the level of gender-sensitivity in the project design by asking questions such as: Did women have separate community meetings in order to express their ideas freely regarding project design? Were women's needs taken into account? To what extent were women involved in the project? These types of questions at the levels of access and participation allowed us to be able to assess progress towards gender equality without tracking any results.

Development initiatives, especially those involving intricate social relationships, can often lead to unforeseen results or secondary impacts. Therefore, we have found that results tracking needs to also assume some elements of "goal-free" analysis and open consideration of effects and impacts,

which go beyond narrow prescribed constraints of projects and programs.

These efforts to support establishing realistic yet ambitious goals with regard to how projects, programs, and organizations take on gender equality require evaluators to engage with those entities providing the funding – and have certain expectations for gender equality results—and those entities implementing programming in a particular context. There are often large gaps in terms of expectations between donors and implementers and evaluations provide a good opportunity to encourage dialogue on these issues.

An example of a project where results were not yet seen due to the agency's stage of development was the evaluation of a program aiming to support several countries in the Caribbean in the strengthening of their social infrastructure. We were asked to find out what gender equality results had been achieved by the program. Although the program did not include any specific activities to enhance gender equality, it did require the project staff to adopt some basic gender-sensitive measures such as ensuring gender balance in the meetings to be held with the project beneficiaries or ensuring that both men and women would benefit from a specific project. We looked at the project implementation process to assess the agency's progress towards gender equality rather than desperately looking for changes in gender relations in the different countries where the program was implemented. We also jointly identified with the management and project staff what were realistic yet sufficiently ambitious project goals. In other words, we tried to identify what types of results the program could realistically take credit for.

Challenge 5: Lack of specific methodological or applicable frameworks for evaluating gender

The lack of established methodological or applicable frameworks specific to the evaluation of gender leaves consultants without a pre-established structure to follow during their evaluation work. Although this may make gender-oriented evaluation tasks more complicated, it also provides consultants with greater flexibility to respond to specific project needs. Furthermore, it forces evaluators to open their eyes and minds to all possibilities, and it encourages them to strive for deep understanding of what is actually going on rather than going through mechanistic rituals that can result when frameworks are overly prescriptive.

Even though formal gender evaluation frameworks are scarce, gender evaluators have used a set of different approaches to understand the role of gender in development. These approaches have either intentionally or unintentionally guided the evaluation of gender. The three major approaches are the Rights-Based/Equity Approach, the Feminist Approach, and the Ethnographic Approach. The Rights-Based/Equity Approach is one of the most commonly used approaches today. It is structured around the notion that men and women should have equal access and representation. This approach is typically quantitative, looking at aspects such as the number of men and women represented or the number of gender focal points within an organization, etc. The Feminist Approach advocates for changes in the power structure between women and men. It is often agenda-driven and can alienate or intimidate non-Feminists. Given that it often pushes a social agenda, it is of questionable applicability to evaluators who are engaged to

present a balanced set of values. The ethnographic approach is structured around the notion that gender relations are subject to their social and cultural context and that therefore, cultural aspects should be taken into consideration when conducting gender evaluations.

The majority of Universalialia's work follows the Rights-Based/Equity Approach because most of our clients use this approach. However, we are trying to move towards the more comprehensive Ethnographic Approach, which looks at social and cultural factors that influence or determine gender relations. We believe that the Ethnographic Approach facilitates a deeper understanding of gender relations and how they are constructed. The deep understanding that can result is also helpful in better assessing programs and organizations and their performance with respect to gender equality.

Challenge 6: Including gender in the evaluation terms of reference

We believe that the best way to include gender in the Evaluation Terms of Reference is to have ongoing dialogue and discussion with the client and other stakeholders to ensure shared understanding and to provide opportunities to explore key issues and questions. Often, dialogue can help clarify terminology, expectations, and possible approaches to evaluating gender equality. Through open conversation, pertinent gender equality issues can be highlighted or the focus of the evaluation can be slightly shifted as issues are explored in more detail. We have found that active dialogue can be helpful to both the client and the evaluator.

Conclusion

This chapter has aimed to outline some of our experiences with evaluating gender, reflect upon some common challenges, and outline a few ways in which we have tried to mitigate them.

A lot remains to be learned and explored in the relation of evaluation and gender equality issues, and we hope that we will have many opportunities to continue learning with and from other evaluation practitioners, clients, and their stakeholders. Active and ongoing dialogue, where mitigation strategies and lessons learned are openly shared, is invaluable to this kind of learning. As part of a step forward, Universalia has recently initiated a professional dialogue on gender equality with its colleagues from outside the firm. We noted that we do in fact share many similar experiences and that sharing these as well as somewhat successful strategies to mitigate challenges can help improve everyone's performance. We believe that further discussion and exchange of ideas not only among evaluation professionals, but also more broadly with clients, gender equality advocates, and others will play an integral role in our ability to effectively evaluate gender issues.

We enjoy performing gender-related evaluations because we share the belief that equality issues are, and should be, at the very core of all development work. However, we are also aware of the complexities of addressing these issues. As we have demonstrated above, there are many challenges when it comes to effectively tracking and demonstrating progress towards gender equality. While we have in some cases been successful in mitigating these challenges, many issues are far from being resolved, such as the question of how to track progress of gender equality initiatives in a realistic and

meaningful way. Yet each new evaluation provides us with another opportunity to learn and improve. Most important is the process of working with clients and their stakeholders to identify issues that need to be addresses and to discuss possible methods for resolving those issues. As Universalialia moves past this 30-year milestone, we look forward to continuing our work in all areas, including that of gender.

CHAPTER TEN

How M&E is 'Perceived' and 'Applied' in the MENA Region

Doha Abdelhamid

M&E of development efforts are intertwined with the development efforts themselves, and often provoke similar reactions to the reactions people have to development trends such as globalization. Globalization cannot be justified on its own right; it can be justified when it fosters development by the people and for them. Anti-globalizers are keen mostly on portraying the weaknesses of globalization; while pro-globalizers tend to emphasize the contrary.¹⁵⁰ The ongoing debate on the global front needs to pause and reflect on the impact of globalization on the "winners" and "losers" in the ongoing process of collapsing walls and borders leading to fast-track mobility of goods, services, benefits, costs and, notably, risks to sustainable development.

Development is all about "people." Countries that have been able to transmit the macro-gains of globalization to the micro-aspirations of local citizens have realized the promise of globalization. Some other countries have retained a classical view and have made their case for globalization based on the measurement of "growth indicators" rather

¹⁵⁰ Stiglitz, Joseph (June 2002). *Globalization and Its Discontents*, W. W. Norton and Company, USA.

than on indicators related to the impact of development policy, program and project interventions on people. Democratic participation, improved governance, and public accountability are important aspects of development impact for grassroots citizens, and in modern M&E they need to be measured and understood.¹⁵¹

Despite its many riches in general, the MENA region continues to be at the dawn of development and of measurement of development impact. This is not surprising given that many aspects of governance within the region lag behind established practices in other regions. In much of the MENA Region, the mode of affairs is based on an autocratic/authoritarian/ paternalistic history. The voice of people is inhibited largely and the security arm of governments is heavy handed over parliamentary auditing tools, as is the influence of political parties. It is not that the people have no clue about their needs -- they know them but are constrained by the surrounding environment. The numerous attempts to globalize within MENA and modernize state structures and move towards extroversion are often reversed as a result of strong cultural factors.¹⁵² Consequently, genuine interest in the cause of development monitoring and evaluation is yet to come. This global public good component has to be considered and enforced as part of the aspired multi-tiered global governance framework.

¹⁵¹ Abdelhamid, Doha (16-17 March 2008). 'From Policies to Development Impact: A Modern Perspective to Economic and Fiscal Policies in Egypt,' keynote conference paper, Center for Political Research and Studies, Public Policies Conference: Between Development Necessity and Effectuation of Citizenship, Cairo University.

¹⁵² Abdelhamid, Doha (2005). Regulatory Competition in An Open Market World: Experiences and Lessons Learnt from the Developed and Developing Financial World, Winner of the Al Ahrum Foundation Regional Award for Best Intellectual Scientific Contribution—2008.

This chapter diagnoses the status quo of development monitoring and evaluation (M&E) in the MENA region, and offers a normative action agenda to link the Region to the relevant globally managed governance system.

The Development and Development Evaluation Gap in the MENA Region

Development thinking in the recent past has been strongly influenced by the importance of adopting a 'people-centered' approach. This adds to the growing consensus that development is all about enhancing individual freedoms, expanding human capabilities, widening choices, and assuring citizens of their basic human rights, while recognizing the intimate interconnection between the fulfillment of social, economic, political, and civil rights.¹⁵³

However, in the MENA region, development is reduced and measured in official announcements mainly in terms of economic growth and wealth accumulation with disregard to genuine development challenges, such as growing poverty, mal-distribution of income, new public sector management, a consensus on a roadmap incorporating measurable reform, etc.¹⁵⁴

This is reflected in the findings reported at a regional conference held in March 2008 in Bibliotheca Alexandrina, Alexandria. The Conference represented the Annual Assem-

¹⁵³ Segone, Marco, and Ada Ocampo (eds., November 2006). "Creating and Developing Evaluation Organizations,' IOCE, UNICEF & CDEPD.

¹⁵⁴ Abdelhamid, Doha (9-11 April 2008). 'How M&E is Perceived and Applied in the MENA Region?,' keynote address to the 'Outcome Mapping Users in MENA Workshop,' the International Development Research Center (IDRC) and the American University in Beirut (AUB)—Environment & Sustainable Development Unit (ESDU), Cairo.

bly of the Arab Reform Forum in response to calls by civil society and governments in the region in what was called the assault, but was formally known as the Broader/ Greater Middle East Project, announced during the G-8 Meeting in Sea Island in 2003. The reason for using the term 'assault' is that people of the MENA region perceived this Project as a dissipation of the Arab identity and language, despite its cultural richness and diversity. It was, and is still thought that this Project implied that Arab nations would be en-grossed in a terrorist-war zone in the aftermath of 9/11. This led Arab leaders and civil society to launch the Arab Reform Forum and to initiate their own reforms from within the region.

The March 2008 conference considered the first tracking report of the 'perception' of the Arab intelligencia (a sample of 494 who were exhaustively surveyed) on economic, political, social, and cultural reforms in the Arab region.¹⁵⁵ A composite index generated data for tracking development in the Arab region for the first time. The report concluded that the overall Arab Reform Index achieved 52.2 percent of the aspired reforms, broken down as follows: Cultural reforms assumed prime position at a 57.3 percent progress rate; political reforms at 51.2 percent; economic reforms at 47.2 percent; and finally social reforms (relating to income distribution, poverty, and social safety nets) at 46.9 percent.

A deeper look classifies Arab countries into three categories, similar to those in the acclaimed Arab Human Development Reports (AHDRs): 1) Oil-rich countries that achieved above-average progress -- United Arab Emirates (UAE), Bahrain, Kingdom of Saudi Arabia (KSA), Oman, Qatar, and Kuwait;

¹⁵⁵ Bibliotheca Alexandrina (March 2008). The Arab Reform Forum Monitor, Part I, the Report, Bibliotheca Alexandrina, Alexandria.

2) Mid-rated countries, which represented the majority of Arab nations, including Egypt, Lebanon, Iraq, Jordan, Palestine, Syria, Morocco, Algeria, Tunisia, and Libya; and finally, 3) The below average human development countries in the Arab world, namely Somalia, Sudan, Mauritania, Djibouti and Yemen.

Not only did the Alexandria 2008 Arab Reform Report show development deficiencies in the region, but the AHDRs issued since 1999 and the Millennium Development Goals Reports (MDGRs) to date show repeated shortfalls in the region in human development on the one hand; and democratic governance, accountability and transparency, on the other hand. The reports attached to these deficiencies a feeble capacity for development monitoring and evaluation.

Of course, the AHDRs and MDGRs affirmed knowledge, technology and digital divides, gender, future generations & youth, climate change & water issues all leading to possibilities of the non-attainment of the MDGs by 2015. This comes in stark contrast to the resource abundance the region is endowed with, namely, a youthful labour force, large reserves of oil, ample agricultural crops and numerous tourist attractions. The Reports harped on the lost development opportunities faced by the Arabs and the need for more accountable governments.

The Paris Declaration in 2005¹⁵⁶ ratified by 116 developed and developing nations, 26 donors, and 14 civil society organizations drew parallel conclusions for the Middle East Arab region. Of the signatories, nine countries (approximately 8 percent) were Arab (Jordan, Kuwait, Morocco, Tunisia, Egypt, KSA, Sudan, Syria and Yemen); a single

¹⁵⁶ Segone, Marco (ed., January 2008). *Bridging the Gap*, UNICEF, IDEAS, WB, DevInfo, MICS, Geneva.

Arab donor (the Arab Bank for Economic Development); and not a single Arab civil society organization.

Despite the grim outlook above, a positive conclusion can still be drawn. The mere fact that Arab countries represented 8 percent of the total signatory countries contrasts with the fact that those nine countries represent 41 percent of the total 22 Arab countries. And, because countries signed the Paris Declaration agreed voluntarily to be subjected to 12 indicators of progress on aid effectiveness to be achieved by the year 2010, they were grouped into five major categories; i.e. ownership of country development strategies; alignment of donors to partner country public financial and procurement systems; harmonizing aid disbursement plans on development program-bases; managing for results through the institution of Results-Based Management (RBM) systems and tracking performance measures transparently; and both parties mutually accountable by performing joint evaluation missions to assess development progress and aid effectiveness.¹⁵⁷

Therefore, it can be assumed that the nine Arab signatory countries already have M&E systems in place, or 'thought/perceived' that they had them.

Then came 2006, when 34 signatory countries around the globe subjected themselves voluntarily to the first baseline study of the Paris Declaration; of the 34, only two Arab countries came forward to engage in the survey (approximately 6 percent), Egypt and Yemen. Both scored 'Ds', the lowest rating of all on 'performance assessment frameworks,' while many other African countries rated better positions (e.g. Kenya, Ghana, Ethiopia, which were rated at a

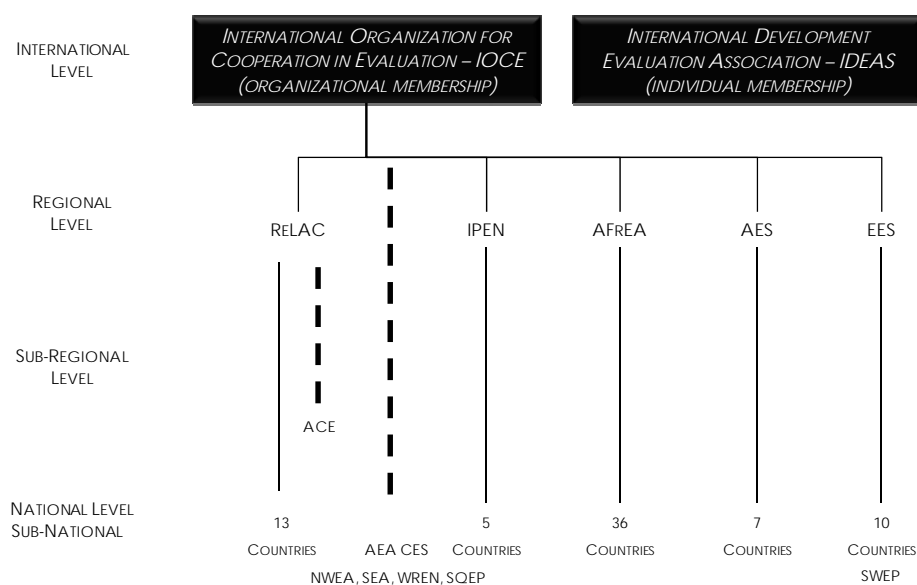
¹⁵⁷ Organization for Economic Cooperation and Development (OECD). Paris Declaration, internet.

'C;' and Tanzania and Uganda that received the highest rating among those surveyed, i.e. 'Bs').

The M&E Network Architecture

Despite this poor performance, things are beginning to change as evaluation capacity is being developed around the world. The overall picture of the evaluation associations and networks architecture is as follows:

EVALUATION ASSOCIATIONS AND NETWORKS¹⁵⁸



Over the last years, the number of national, regional, and international evaluation associations evolved from half a dozen in 1997 to more than 50 in 2007. Most of these new

¹⁵⁸ Source: Quesnel, 2006

organizations are located outside Western Europe and North America, which is to say mainly in developing nations.

Two global organizations: the International Organization for Cooperation in Evaluation (IOCE) and the International Development Evaluation Association (IDEAS) were launched in 2003, with organizational members in the former and individual-practitioner evaluation members in the latter.

The following tier includes regional evaluation organizations such as the Relac (Red de evaluacion de America Latina y el Caribe—13 countries); IPEN (International Program Evaluation Network [Russia & newly independent states—evalnet]—5 countries); Afrea (African Evaluation Association—26 countries); AES (Australasian Evaluation Society—7 Australian-Asian countries); and EES (European Evaluation Association—10 countries); then follows the national and local tiers, respectively.

The observation drawn from this analysis is: NO Arab society was included in the IOCE roster; an Israeli Association for Program Evaluation (IAPE) from the Middle East and North Africa (MENA) region is listed, the same for an organization named the Egyptian Evaluation Society, that had a contact that is inoperative, does not possess an internet website or a known mailing address.

One can easily draw the following conclusion from the just-mentioned evidence: The Arab people are not only deprived from democratic governance, accountability, transparency, human development and lost opportunities; they are also deprived of tracking their development progress worse from what is commonly known on the global front as ECD (Evaluation Capacity Development) to a large extent.

A pressing question that comes to mind immediately is: How would citizens know that they have gained the fruits promised by their elected political party candidates, unless their steps are measured? How else can people be positive about the route treaded and future directions?

Who Are the ECD Target Groups?

In order to do proper ECD targeting in the Arab region, which is currently very deficient in many respects, an ECD strategy has to be developed and target groups have to be identified clearly.

Literature surveys tend to emphasize that the only stakeholder reported at length seems to be 'governments.'¹⁵⁹

An ECD strategy should define the ways and means to strengthen M&E skills of each of the following groups:

- Agencies that commission and fund evaluations. These include development partners, foundations, government budget and funding agencies, and national and international non-governmental organization (NGOs)
- Evaluation practitioners who design, implement, analyze and disseminate evaluations. These include evaluation units of line ministries, planning and finance ministries, national and international NGOs, foundations and development partners, evaluation consultants, and university research groups

¹⁵⁹ DBSA, AfDB, & WB/OED (25-29 September 2000). M&E Capacity Development in Africa, Johannesburg.

- Evaluation users. These include government, development partners and civil society organizations that use the results of evaluation to help formulate policies, allocate resources, and design and implement programs and projects
- Groups affected by the programs being evaluated. These include community service organizations, farmers' & trade associations, business groups, trade & workers' unions and many other groups affected directly or indirectly by the programs and policies being evaluated
- Public opinion. This includes broad categories such as the general public, the academic community and civil society.¹⁶⁰

Further retort on the specific M&E skills required by each group can be elaborated, yet the point that has to be made is the need to 'look at' and 'perceive' stakeholders differently than the way the literature normally does. Even more, 'evaluator-practitioners' themselves have to denounce the attitude of being self-conscious of their very being to consider any other group out of their circle as 'their stakeholders.' This should be replaced by a perspective whereby the 'evaluator-practitioners' stay as is while their effective stakeholders have in-built M&E or ECD capacity as well. If this is managed well, Arabs in the general sense of the word will speak and practice the evaluation language of the world around.

¹⁶⁰ See Segone and Ocampo November 2006.

Even Development Partners Err

Despite scarcity, a meticulous M&E literature reviewer would come across two donor reports, specifically produced by the United Nations Development Programme (UNDP), in the form of workshop proceedings held in Syria in 2005 and Jordan in 2007¹⁶¹, to arrive at the conclusion that they comprised the main represented view of the United Nations (UN) staffers in the Regional Bureau of Arab States (RBAS) New York and in the Arab region voicing loudly the deficiencies existing in the UN evaluation system. Points were made on the lack of capacity of in-country office staff relating to the M&E science, practice, terminology, and methodologies required for outcome and impact assessments. They also found that, at many instances, program designs are de-fragmented due to the lack of robust logical frameworks linking policies to programs and projects the way known as the "policy evaluation cycle"¹⁶² around the globe, insufficient knowledge of appropriate outcome and impact indicators; the same for lack of accountability of senior program country officers to headquarters' management.

Surprisingly, in the back of the minds of many Arabs, as much as the rest of the world, UNDP was 'thought' and 'perceived' to be the leading organization in the institution and 'application' of M&E systems; and how can this be negated while global, regional and national human and millennium development reports continue to be generated

¹⁶¹ UNDP (December 2007). *Achieving Results, Evaluation of Results-Based Management at UNDP*.

¹⁶² Abdelhamid, Doha (2006). 'Use of Evaluation in the Policy Cycle,' in depth workshop training materials, World Bank, International Finance Corporation, and Carleton University, Ottawa.

punctually and are, in fact, used as a yardstick reference in the Arab region, and others, for periodic development metrics. Even this measure has casted scepticisms to the effect that once evaluated, those national, regional and global reports were considered "output" measures of development, except perhaps some recent reports starting in 2002 that came to generate more "evaluative knowledge."

The courage shown by the UNDP in identifying its M&E weaknesses and in acting to develop its first evaluation strategy in 2007 has to be encouraged and acknowledged. Self-criticism and assessment stands to be a useful tool so long as it is taken constructively to recourse en route.

Thus, the faint evidence on the presence of coordinated and integrated ECD effort and M&E practice in the Arab region (except sparingly in some cases that came to the author's personal knowledge such as in Morocco, Bahrain, Kuwait, Jordan and Egypt) this can be attributed to the mere recognition, again, that accountability and governance mechanisms remain a taboo in many Arab societies with some recent progress; and that sustained M&E cultures do not sufficiently exist except in a patchy manner through the vehicle of donor-funded projects mainly, rather than through well-designed and coordinated programs, with genuine stakeholder participation and mapping.

What Do Arabs Have To Do?

A synthesis of recommendations elicited from a series of workshops/conference reports generated by the donor community and especially the World Bank (WB) voiced the needs of government officials, civil society, research institutes, and local universities in Africa and Arab nations in

developing M&E capacities, seem worthy of careful scrutiny by Arabs and others interested in this region that is to say, the donor community, Universalia, the research community or any other:

- The main precondition for the development of a national evaluation system is country 'demand'—an evaluation system cannot be effectively forced on an unwilling government (the case of Egypt¹⁶³ is evidence). There are particular risks if the impetus for an evaluation system is donor-driven; this is not to say that donors cannot take the lead in 'selling' or 'promoting' the merits of evaluation systems to countries, but that unless and until countries accept the strength of such arguments, or reach their own conclusions that evaluation has much to offer, an evaluation system is unlikely to be sustainable.
- Two building blocks for effective demand are the sensitization of key stakeholders to the need for and benefits from evaluation, and the building of awareness of suitable techniques and approaches
- Experience informs that the main barriers to developing evaluation systems in developing countries have been: lack of genuine demand and ownership in countries; lack of a culture of accountability, often related to issues of ethics and corruption; lack of evaluation, accounting, or sufficient auditing skills—there is a need to develop the supply of these governance-enhancing skills and systems to match demand as and when it grows; poor quality of financial

¹⁶³ In her earlier capacity as fiscal reform programs manager, the author acknowledges the program of performance-based budgeting monitoring and evaluation capacity development in partnership with UNDP, WB and local expertise (January 2001-July 2004).

and other performance information, and of accounting/auditing standards and systems; lack of evaluation feedback mechanisms into decision-making processes; and the need for the efforts to develop evaluation systems to have some minimum critical mass to be sustainable.

- Another lesson of experience is that the development of an evaluation system should not be viewed as a stand-alone activity if the institutional framework and incentives do not support it. If the framework and incentives (carrots and sticks)¹⁶⁴ are insufficient then this is a strong argument for ensuring that the development of an evaluation system is part of a broader initiative to develop governance; this approach recognizes synergies between performance measurement/evaluation and performance management.

Catalytic Factors for ECD Success

Developing evaluation capacity implies a need for change, and as is the case in any part of the world, major change to the dominant paradigm are met with fierce resistance. In many respects, the resistance is understandable as there are winners and losers for any paradigm shift. In my different political, government, private sector and civil society capacities, a group of like-minded persons and I continue to hammer on the need for reforms, and those that go beyond reform for the sake of reform. Success will require continued commitment to the cause of M&E and continuing writing

¹⁶⁴ Bemelmans-Vidéc- Marie-Louise, Ray Rist, and Evert Vedung (eds., 2003). Carrots, Sticks and Sermons, Transaction Publishers, USA.

and speeches to many people in different sectors on the need for ECD. I have personally experienced an exponentially growing interest in the field from an assortment of stakeholders.

Experience has also revealed that a number of success factors exist for the development of an evaluation system, and that a system should be pursued only if a number of “catalytic” factors already exist or if there are reasonable prospects for developing or creating them.

One factor is the powerful role which a 'champion' ministry or agency can have in supporting, encouraging and pushing the development of an evaluation system. This requires a lead agency with power and influence, such as the Cabinet of Ministers machinery, a Finance or Planning ministry (as in Morocco, Kuwait, Bahrain and Egypt), a national audit office, or a civil society organization, etc. Added to this is the development of a core group of supporting ministries to walk the talk. The existence of an explicit and high profile country evaluation strategy can also be effective in selling the message, as can the support of individual ministers or the Cabinet as a whole—particularly through a sustained legislation, ministerial or presidential decree.

Sustained government (or simply 'champions') commitment is also important (The case of Egypt is evidence of this -- when commitment vanished, the system collapsed).¹⁶⁵ An evaluation system cannot be developed overnight; indeed experience indicates that it can take at least a decade at the whole-of-government level to embed such a system in a sustainable manner – i.e. to gain an understanding and belief

¹⁶⁵ Abdelhamid, Doha (29-30 April 2004). 'Painful yet Rewarding: Paving the Road to Results-Oriented Budgeting in Egypt,' Conference Paper, Parliamentary-Government Policy Dialogue on Reform and Fiscal Issues, World Bank, Rabat.

of the system's merits, to develop necessary skills within responsible organizations, and to set up civil service structures and systems to implement evaluation findings. The counterpart of the need for sustained government support is the need for sustained support from development partners.

A whole-of-government approach (as the ones applied in Indonesia, Chile, Canada, and Australia) has an advantage in achieving momentum and helping to ensure that laggard ministries endeavour to keep up with leading ministries. A whole-of-government approach might be especially feasible if a major series of reforms is being contemplated, such as major changes to public expenditure management (i.e. budgetary processes and decision making) in response to budgetary imperatives (Kuwait and Bahrain applied a whole-of-government approach within a Medium-Term Expenditure Framework (MTEF)/Public-Expenditure Tracking System (PETS) and after piloting)¹⁶⁶. However, a whole of government approach might be unrealistic in some developing countries -- it can be difficult to achieve due to resource or local capacity constraints, or both. In many countries a more achievable and modest approach might be to start with an initial focus on ongoing performance monitoring, possibly in particular sectors or ministries (Kuwait started with education in 2006, health in 2007, then moved to whole-of-government in 2008) in order to create a demonstration effect. Egypt adopted a pilot approach¹⁶⁷, and then sought to spread the approach to other sectors/ministries and to include other performance measurement tools (such as evaluation) as opportunities presented

¹⁶⁶ Abdelhamid, Doha, and Sahar El Tohamy (2004). 'Fiscal Autonomy and Public Accountability,' Egypt Human Development Report, the United Nations Development Program (UNDP) and the Institute of National Planning (INP) of the Ministry of Planning (MOP), pp. 53-68.

¹⁶⁷ See Abdelhamid 29-30 April 2004.

themselves. This type of sequencing implicitly depends on perception of satisfactory benefits vis-à-vis costs across time sufficient to generate demand for answers about program outcomes and causality and these are the types of questions that only summative/impact program evaluations can fully answer.

Next, there is a need to tailor the approach for developing evaluation capacity to the circumstances and opportunities available to and in each country or context, so no one-size-fits-all.

The final lesson is that incentives are crucial to ensuring both that an evaluation system is developed and that evaluation findings are actually used. In working with national governments and elsewhere, it is important for those providing technical assistance to understand the incentive frameworks in the country. This would typically involve conducting an institutional diagnosis as a precursor to provision of advice, and to ensuring that this work includes a close dialogue with beneficiaries. Zimbabwe followed a similar approach, and it worked well. This approach should also examine the extent of evaluation capacity in a government or organization—it is often the case that governments, specifically (unlike private sector firms who are always under cut-throat competition), over-estimate the extent and quality of their evaluation capacity.

An assessment of who would lose if information is disseminated and who would be held accountable is an important consideration for the evaluator. He/she will have to postulate the effect on accountability and 'behavioural changes,' which are a key area in OM.

This leads to the issue of who should be responsible for measuring performance -- impartial outsiders, or expert

insiders. The former approach stresses objectivity and independence, while the latter stresses inside expert knowledge and ownership of the evaluation results, and this in turn is likely to encourage learning by managers and their staff. Therefore, evaluators should always emphasize the importance of learning from improved performance and emphasize as well accountability leading to better governance and transparency.¹⁶⁸

Conclusion

As an Arab, the author puts herself in place of all Arab people and lists a set of points on what she thinks is wanted from the global and donor community to enhance appreciation of the intrinsic value, perception and practice of M&E:

- Arabs need development partner agencies to view the provision of ECD training and technical assistance as part of a sustained, ongoing partnership and dialogue with Arab governments, private sector, civil society, and citizens. The role of development partners cannot be neglected—they can build conducive environments or hinder the development of evaluation capacity through such partnerships.
- Arabs need development partners around the globe to work in partnership with national governments, consultants, civil societies and citizens in enhancing accountability through developing and implementing democratic governance programs with support components to parliaments, political parties, media;

¹⁶⁸ Mackay, Keith (May 1998). ECD: Lessons from National Experience, ECD Working Paper Series No. 1, World Bank, Operations and Evaluation Department, Washington DC.

and open a dialogue with governments on passing data dissemination legislation and M&E action plans with funding support from program budgets.

- Arabs need development agencies to offer technical assistance—through their advice and sharing of lessons and good practice, as much as the provision of funds for methodological development, dissemination, and training in centers of excellence, distance, or CD learning tools. This will help in capturing credibility for the trained and perhaps encourage the development of regional and national professional networking organizations. Donors can also help to build national and regional evaluation systems and literature in Arabic for ease of use and access. Their support will offer governments assurances and confidence on the robustness of reforms adopted.
- Conversely, donors can hinder by excessive, conflicting or multiple donor requirements for evaluation. There is a particular danger that scarce country evaluation capacity will be diverted to satisfy donor requirements for the evaluation of development assistance activities, while neglecting the main reform area undertaken. A priority is to build a grand evaluation-capacity base to sustain and support M&E skill shortages in the Arab region.
- Arabs need their development partners to introduce new lending instruments, which provide multi-year soft funding for TA activities tailored to meeting country needs for long-term commitment and support in this area.

- Arabs need development partners to coordinate donor funding pools for joint evaluations by governments and donors that can be helpful in reducing costs and reducing duplications in evaluations.
- Sharing collective knowledge repositories in the field of developing ECD for closer donor collaboration and partnership in this area is important, as the Arab region lacks a documented track record of evaluation cases. This could take place via: the sharing of existing case studies, which are invaluable learning tools on country contexts; identification and analysis of existing good practice or promising practice countries; development of approaches undertaking country diagnoses and identifying their potential to put evaluation systems in place; and sponsorship or other support for regional collaboration including the sharing of experience via virtual discussion group networks, conferences, and/or workshops as the one we are in right now.

Finally, M&E is all about building a performance culture in the mindset which takes care of the 'perception-part' of it, and in the daily 'practice' that will infiltrate sooner or later to the qualities for and by all the people in the Arab region if persevering and collaborative work is put in place for the development future of the region in the light of the hegemonic power shifts in the globe and the new political economy for managing global governance prospects worldwide. Development M&E is indeed a public good of value at no lesser importance than managing global governance itself.

SECTION IV

Conclusion



CHAPTER ELEVEN

Concluding Thoughts

*Marie-Hélène Adrien, Charles Lusthaus,
and Gary Anderson*

For thirty years, Universalia has been engaged in conducting evaluations, training evaluators, building evaluation capacity, creating evaluation frameworks, contributing to the evaluation literature, coaching evaluators, advocating for better evaluation practice, and building a global evaluation community. The last 30 years was a period of time in which the field of evaluation came of age. With our Canadian and global colleagues, we were pioneers in providing better evidence to a large group of national and international investors who were trying to reduce global poverty and improve the lives of billions of people.

The focus of our evaluations was aimed at providing the best possible information to help decision-makers determine where to focus with respect to future projects, programs, organizations, institutions and policies. It was and continues to be a humbling experience and a role that the Universalia team and the international evaluation community take very seriously. While in some ways the pioneering stage continues, in most of the world evaluation is only now becoming a serious and valued endeavour. Even in North America, evaluation is still not widely acknowledged as a discipline, though as Michael Scriven concludes, the situation is beginning to change.

*While evaluation has been practiced for many years, it is only now developing into a discipline. In this way, evaluation resembles technology, which existed for thousands of years before there was any substantive discussion of its nature, its logic, its fundamental differences from science, and the details of its distinctive methods and thought. In recent years, we have begun to see more discussions within the field about evaluation-specific methodology. We are moving toward the general acceptance of evaluation as a discipline, but there is still a long way to go.*¹⁶⁹

Ian Green, looking back at two decades of evaluation practice, reminds us that within the past 20 years an impressive array of national evaluation organizations has emerged all over the world:

*... an explosion of interest in the utilization of evaluation techniques in Europe and other parts of the world, as reflected by the birth of new evaluation societies. The European Evaluation Society was formed in 1994, and in the same year, the United Kingdom Evaluation Society appeared. Evaluation societies were started in Australia in 1991, France in 1996, Germany, Italy and Switzerland in 1997, and Malaysia in 1999. Recently, evaluation societies have been formed in Russia, Belgium, Ethiopia, Finland, Ghana, Israel, and Kenya. No doubt, their members will have the same innocent enthusiasm that Canadian and American evaluators started out with, but it is hoped that these new societies can take advantage of some of the lessons learned by their North American counterparts.*¹⁷⁰

¹⁶⁹ Michael Scriven (2003-2004). The Evaluation Exchange Harvard Graduate School of Education, Vol. IX number 4, winter 2003-2004.

¹⁷⁰ Ian Green Lessons Learned from Two Decades of Program Evaluation in Canada <http://www.yorku.ca/igreene/progeval.html>

Finally, Jean Quesnel conceptualizes the institutional evaluation architecture, which incorporates the private sector, governments, institutions, NGOs, and various networks of each of these organizational types¹⁷¹. Local, national, regional, and international organizations are growing and developing into a community of practitioners, scholars and “users” all aimed at improving the usefulness of evaluation feedback to key decision makers. Universalia is proud of its small, but we hope meaningful contribution to both the substance of the field as well as the development of its global architecture.

However, that was the past. At this stage in Universalia’s life, we are ready for the next 30 years. Where is the “discipline” of evaluation going? What challenges lie ahead?

As we think about the future we have identified ten trends that we believe have generally taken hold and will continue to develop and be a force in the evaluation world generally, and the international development evaluation world more particularly. These ten trends are not independent, but form a web of complexity that will continue to make evaluation an intellectually exciting profession:

1. *Improving the quality of evidence provided to decision makers*

When we started Universalia, the “evidence” most often used by decision-makers was expert opinion. Experienced people would provide some information but give their opinion about the quality and future direction of projects and programs. This is becoming less and less acceptable and the present debate about appropriate “evidence” will continue as will the quality of evidence. In general, the drive

¹⁷¹ For a full review of the evaluation infrastructure see Jean Serge Quesnel “the importance of evaluation associations and networks,
http://www.ioce.net/download/other/Quesnel_ImportanceAssoc.pdf

for better evidence will continue spurred by the ongoing 30-year debate between quantitative and qualitative methodologists and those advocating mixed models.

2. *Professionalization of the field*

Over the past 30 years, evaluators were people who labelled themselves “evaluators”. There were no requirements for entering the field, few clients required evaluation qualifications, and many who engaged in the evaluation world did not associate themselves with the body of knowledge, the institutions, the networks, the standards of competing fields such as comprehensive audit. As we are writing this book, the Canadian Evaluation Society has just taken steps in creating a professional evaluation designation. The international evaluation community is discussing a similar initiative. The impact evaluation community likewise is attempting to professionalize its work. In addition to various ways of creating certification, the body of evaluation knowledge is growing, standards are being upgraded and increasingly enforced, evaluation degrees and training programs are being developed by universities and professional bodies. In other words, the field of evaluation will continue to professionalize.

3. *Cultural values and different “ways of knowing” will add complexity to practicing evaluators while enhancing the quality of evaluations*

Evaluation involves the use of social science research tools to answer a set of important applied questions using evidence that is linked to particular judgements: What is good? What is less good? What should continue? What should stop? What requires more investment? What should have less? The evaluator’s techniques are tied to a worldview of who

knows best with respect to what questions we ask, what we mean by “research”, “quality data”, “judgements”, etc.

Thirty years ago, the most valued data tended to be empirical, but over the last two decades, qualitative data has gained respect. Researchers in all fields of social science are beginning to accept that multiple data sources and methods of data collection and analysis are useful. Furthermore, it is now accepted that there are multiple ways of knowing, and that some ways are more conducive to certain questions and situations. We have come a long way from the imperialist notions about research in the past:

“The clash between Western and Indigenous ways of knowing has been epitomized by the “parachuting model” of the Western researcher who drops into the reservation, collects data and leaves never to be heard again...”¹⁷²

This worldview is being challenged and continues to be a major source of concern to the evaluation community not only in North America, but increasingly in Asia, Africa and in South America. It is also the concern of those who argue that women’s ways of knowing are different from men’s ways! In general, this thesis suggests that different communities have different knowledge, value knowledge differently and make judgements based on different ways of knowing and value premises. This methodological relativism is a challenge to those who seek methodological “gold standards” as well as those within the evaluation community and the investors who rely on evaluations to inform their decisions. While these multiple avenues towards truth add complexity, they also offer the promise of better, more thorough, stakeholder-grounded and useful evaluations.

¹⁷² S. Scott CJPE v23 n2 p74 autumn 2008

4. Country-led evaluations will become increasingly the norm

As the global evaluation architecture develops, evaluation capacity will increase around the world. Countries which previously relied on the international community to plan evaluation work will have competent evaluators available locally who can plan and execute evaluations that meet their needs and those of their financial partners. Such country led evaluations will lead to national infrastructure (e.g. university programs, institutes, associations, and publications) that will continue to build the national, regional and international evaluation community. Our challenge will be to figure out how best to be a part of these developments, without imposing our singular “ways of understanding” the discipline of evaluation!

5. Evaluations will focus more on programs, organizations, institutions, and policies

Within the international development evaluation world, there has been a trend to move from project evaluation to program evaluation. This trend will continue and added to it there will be more assessments of organizations, institutions and policies. As we see it, governments need to know more about the overall enabling environment and how it relates to programmatic responses.

This is not simply about programs and projects that are being funded, but also the organizations, institutions, and policies they are supporting. To date there have been relatively few evaluations of these latter three areas of investment. Furthermore, insufficient attention has been paid to the questions, methodologies and decisional criteria (judgements) used in assessing these areas. As we move into the next 30 years, evaluators will be putting more emphasis on these areas.

6. *Supporting the process of managing for results will continue to play an important role in the evaluator's tool kit.*

Everyone wants to see their investments pay greater dividends! No one wants more bureaucratic and technocratic churn! Thus, in many ways the “RBM war” is a silly fight. The main issue is how we learn to provide more and better results! Not more paper! Not more forms! More results!

Evaluators have and will continue to have an important role to play in improving the results of various investments. They also need to play a role in assessing the efficiency of bureaucratic demands. Ex ante evaluations can provide insight into the quality of planning, the analysis of the development problem and the theory of change driving the solution, as well as providing investors with more insight into their level of risk. Development evaluation can play a greater role in improving our ability to deliver.

Useful feedback at the implementation stage is crucial for most activities and evaluators have significant experience in this area—in this sense they should be an indispensable hand to managers who are ultimately responsible for results!

7. *Information technology will continue to force us to change*

We are just entering a period of massive change in the collection, storage, analysis, and treatment of large quantities of information at relatively little cost. The result is that there will be an increasing expectation by governments, civil societies, and policymakers that evaluators will be using this data as part of their assessment process.

The good news is that such information will make our work more robust and useful. The challenge is that many of the monitoring and information systems that we have reviewed are not as useful or as complete as “advertised”. This will change! As Ray Rist points out, we already have some global

models that show that the carefully constructed and appropriately used monitoring systems can help provide managers and investors with the evaluative data they need to make important formative and summative decisions¹⁷³

Evaluators will need to be able to audit and work with these new streams of information to test their reliability, validity, and trustworthiness.

8. Evaluation will increasingly be mandated as a part of normal government operations

Some developed countries and donor agencies now require evaluation as part of the program cycle, or funding renewal process.

This practice will continue to spread and become refined, adding different forms of evaluation at various stages. This offers opportunity to the evaluation community, but also requires responsibility to preserve trust in evaluation practice.

9. The public will become more aware of evaluation and its useful function

Today people know little of evaluation beyond the word itself. What knowledge they do have is often confused by media exposure of parliamentary or congressional committees, auditors general, and the US General Accounting Office, all of which conduct forms of evaluation.

When there are problems, the public demands an “inquiry”, but typically only bureaucrats and politicians anxious to defer action, or remove a situation from a political arena, call for an evaluation. Some governments are beginning to make

¹⁷³ Ray Rist (2003-2004) *the Evaluation Exchange* Harvard Graduate School of Education Vol IX number 4 winter 2003-2004

evaluation reports available on-line, either in their entirety, or in watered-down form. This can be helpful to the public and is useful in exposing the strengths and weaknesses of evaluation standards.

Evaluators have an increasing responsibility to communicate about their work to establish a clear niche and increase public understanding.

10. International evaluation is entering an era of innovation and discovery

The notion of “the external evaluator” is largely passed, now replaced by evaluation teams, and often including internal stakeholders. Evaluation as an event has increasingly been replaced by an evaluation process. Self-evaluation is gaining popularity in certain situations.

We have now seen technology develop so that paper questionnaires are rapidly being replaced by electronic menu-driven forms selected by users according to their demographics and preferences.

We have yet to see evaluations on You-Tube, or even to routinely incorporate media into the evaluation process. We foresee evaluators working internationally with live feed video of data collection in Pakistan or Afghanistan, being fed to multi-national teams of analysts in several countries – this would add efficiency to the process, power to the analysis, and build evaluation capacity of all involved evaluators along the way. An exciting era lies ahead!

Thirty years ago, Universalialia started on a journey that was intended to help a couple of academics link theory and practice through the use and practice of engaging in evaluation. Evaluation was the means to help make the world—just a little bit better! We were idealistic and naïve! Thirty years later, the thirty-five staff and countless associates of Universalialia look forward to continuing to contribute to improving the quality of services to people through evaluation! While we are more realistic today, we continue to be idealistic as we look to the future!

CONTRIBUTORS

Doha M. Abdelhamid holds a Doctorate in Financial Economics as well as other degrees and diplomas in economics, computer science and Development Monitoring and Evaluation. Dr. Abdelhamid is currently an Executive Board Director of the International Development Evaluation Association (IDEAS) and serves as Senior Policy Advisor at Egypt's Cabinet of Ministers, Senior Human Resource Advisor at the Egyptian Presidency, Senior Financial Economics and Development Monitoring and Evaluation Expert at the Egyptian Parliament, Senior Advisor to Egypt's Ministry of Local Development, and Senior Gender Advisor to the Arab Labor Organization of the Arab League. She has previously been a Policy Advisor to the Egyptian Cabinet, the Egyptian Ministers of Finance, Economic Development and Planning and Local Government. She has also been Professor at the American University in Cairo, Heriott-Watt University, the Arab Academy of Science and Technology, and the International Program for Development Evaluation Training. | abdelhamiddoha@yahoo.com; dmounir@hotmail.com

Marie-Hélène Adrien, Universalialia's President, holds a Masters and a Ph.D. Degree from McGill University and specializes in Administration and Policy Studies in Education. Originally from Haiti, she grew up in West Africa. Dr. Adrien has led many evaluations since joining Universalialia in 1988, most recently World Bank evaluations of the Global Facility for Disaster Relief and Recovery and the Policy and Human Resource Development Trust Fund, a Caribbean Development Bank Operations Audit of the Basic Needs Trust Fund Offices and a Mid-Term Evaluation of the Basic Needs Trust Fund, as well as evaluations of several UN Organization programs, including the UNESCO Global Monitoring Report – Education for All, and UNIFEM's Trust Fund for the Elimination of Violence Against Women. Dr Adrien has published works on evaluation, including *Enhancing Organizational Performance: A Toolbox for*

Self-assessment, and Organizational Assessment: A Framework for Improving Performance. She is the Past President (2005-2008) of the International Development Evaluation Association (IDEAS). | mhadrien@universalia.com

Gary Anderson graduated from Harvard in Evaluation of Education Curriculum in Natural Sciences, and holds degrees from McGill University in Science, Arts, and Education, where he served as Chair of Administration and Policy Studies in Education for 25 years and as Director of the Office of Research on Educational Policy. He was Co-Founder of Universalia and served as President for over 20 years. Now retired from employment with the firm, he continues to contribute to specific evaluations, and maintains ongoing involvement in Universalia. Most recently, he co-headed an evaluation of UNRWA's Education Program in five countries and territories that host Palestinian refugees. He is author of *Fundamentals of Educational Research*, 2nd edition (1998) and is co-author of various books on organizational assessment. In 2007, he edited *Facilitating Change: Reflections on Six Years of Education Development Programming in Challenging Environments*, an analysis of his six years of work in education in the Balkans. | ganderson@universalia.com

Katherine Garven is an Honours Graduate from McGill University in the field of International Development. She has studied extensively in Latin America and the Middle East and speaks French, English, Spanish, and Arabic fluently. Katherine has experience working on projects in the Middle East, Africa, Latin America, Europe, and North America for UN agencies, national governments, and private enterprises. Her consulting experience is focused primarily on project management and program evaluation in the areas of economic and community development, education, gender issues, and aid effectiveness. She has a strong interest in the management and evaluation of networks and international partnerships and is currently pursuing research in these areas. | kgarven@universalia.com

Ricardo Gomez graduated with a Ph.D. from Cornell University following completion of an M.A. at Université du Québec à Montreal. He is currently Assistant Professor, Information School,

University of Washington, USA. He specializes in research and evaluation of communication technologies in international development contexts. | Rgomez@uw.edu

Silvia Grandi holds a Masters degree in Development Studies from the Institut d'Etudes Politiques in Paris and is currently enrolled in the University of Ottawa Graduate Certificate in Program Evaluation. Prior to joining Universalialia, she worked with Italian and Moroccan NGOs designing projects to support and promote women rights and gender equality in Morocco and the Mediterranean region. Recently, she has been supporting international organizations such as UNIFEM and UNICEF as well as several Canadian Government departments. Her field experience includes work in Africa (Malawi, Senegal, Burundi, and Morocco), Asia (Bangladesh) and Europe (Italy and France). | sgrandi@universalialia.com

Douglas Horton received a B.Sc. and a M.Sc. degree in agricultural economics from the University of Illinois and a Ph.D. degree in economics from Cornell University. Doug is currently an independent consultant specialized in organizational development, innovation studies, and evaluation. From 1975 to 1989, he led the Social Science Department of the International Potato Center, and from 1990 to 2004, he worked at the International Service for National Agricultural Research (ISNAR), headquartered in The Netherlands. At ISNAR, he carried out applied research and provided training and advice to agricultural research managers in the areas of research management and evaluation. Doug has conducted more than 50 evaluations, mostly concerning applied agricultural research and development programs that aim to foster capacity development and pro-poor innovation. His current interests focus on the management and evaluation of partnerships in agricultural research for development. | d.horton@mac.com

Charles Lusthaus was formerly a Professor in the Department of Administration and Policy Studies in Education at McGill University where he was also the founding Director of the Centre for Educational Leadership. He is also a founding partner of Universalialia where he pioneers an approach to institutional and organizational assessment (IOA). Today, his framework is used by a variety of international bodies including CIDA, IDRC, IDB, IUCN, The

Rockefeller Foundation, and ILO. Dr. Lusthaus has assessed a wide range of international agencies including IUCN, ICRC, UNHCR, UNCDF, UNDP, IDB, AfDB, International Trust Funds (WB, UNIFEM), and various Government Ministries. Three of Charles Lusthaus' books explore approaches to evaluating and strengthening institutions and organizations. *Enhancing Organizational Performance* has been one of the International Development Research Centre's (IDRC) best-selling books for the past five years.

| clusthaus@universalia.com

Ronald Mackay was educated at Universities in Aberdeen and Edinburgh and obtained his Ph.D. at l'Université de Montréal. He has held teaching positions at the University of Newcastle-upon-Tyne (UK), University of Bucharest (Romania), Regional Language Centre (RELC) (Singapore), Ontario Institute for Studies in Education (OISE) - University of Toronto and was Senior Fellow, International Service for National Agricultural Research (ISNAR) in 1997. Since retiring as Professor Emeritus (Education) Concordia University, he has worked as an independent consultant in evaluation and organizational diagnosis. He specializes in naturalistic evaluations of development projects (education, agriculture, social development) in the public, private, and voluntary sectors in complex organizational and multicultural settings. He is also an experienced beekeeper specializing in the pollination of commercial avocado production in Chile. | mackay.ronald@gmail.com

Elisabetta Micaro's educational background focuses on international development and cooperation. She obtained a Master's Degree in international studies from Laval University (Quebec) in 2007 and a certificate in international cooperation from the Institute for International Policy Studies (Milan) in 2003. Her undergraduate and graduate studies took place in Italy, Belgium, and Canada. Elisabetta Micaro's main areas of focus in evaluation at Universalia are the evaluation of gender-specific and gender-related development initiatives and the integration of the human rights-based approach and gender perspective into evaluation.

| emicaro@universalia.com

Peter Morgan graduated with a BA in History and a MA in Public Administration from the University of Toronto and an MPA from

Harvard in 1988. He worked with CIDA until 1991 as Program Director for Indonesia and subsequently became responsible for a staff group charged with coordinating management policies and delivery strategies for CIDA's bilateral programs in Asia, Africa, and Latin America. Since becoming an independent consultant, Peter Morgan has pursued his passion for capacity development and institutional development projects by studying and supporting such organizations as the Aga Khan Foundation, UNDP, and the Global Environment Facility. | Morganpj@aol.com

Shaun Pather holds a Doctor Technologiae (IT) from Cape Peninsula University of Technology and a Masters Degree in Public Administration from the University of KwaZulu Natal, South Africa. He was a Fulbright Visiting Scholar, Information School, University of Washington, USA. He is currently Associate Professor, Informatics & Design, Cape Peninsula University of Technology, South Africa and specializes in the evaluation of ICTs in community development contexts and the evaluation of the effectiveness of Information Systems within e-Commerce and e-Government environments. | pthers@cput.ac.za

Ramón Pérez Gil graduated from the School of Sciences at UNAM (Mexico's National University), and has a Master's of Science in Wildlife and Natural Resources Management from the University of Michigan. Ramón Pérez Gil is currently Program Director of the Fundación Gonzalo Río Arronte IAP in Mexico. He is also President of FAUNAM A.C. and Director of PG7 Consultores, S.C. His work focuses on research and technical assistance in the areas of conservation and the sustainable use of wildlife. He has taught in several universities and has been involved in the planning, managing, monitoring and evaluation of wildlife and natural resource projects for organizations such as the World Wildlife Fund, IUCN, USAID, Inter-American Development Bank, World Bank and Birdlife International. | pg7faunam@yahoo.com.mx

Katrina Rojas has a Master's in Public and International Affairs from the Woodrow Wilson School at Princeton University. Between 1994 and 2001, she lived in Costa Rica and consulted on organizational development issues with civil society organizations involved in children's rights, human rights, women's rights and gender

equality, health and environmental issues, popular education, and community development in Central America. She is currently Senior Project Manager and has been a Partner at Universalia for five years. Her most recent work has focused on organizational assessment and the evaluation of global trust-funded programs. In addition, she has a special interest in the evaluation of capacity development and policy influence. The promotion of gender equality and women's human rights has been another area of professional focus preceding her involvement in evaluation work, and this interest now continues in the context of evaluation. | krojas@universalia.com

Ken Stephenson has a Master's of Business Administration (Finance Concentration) and Graduate Certificate in Program Evaluation from McGill University. In his past five years at Universalia, Ken has helped develop M&E systems and has conducted monitoring, evaluation and other types of studies for clients around the world, including UNIFEM, the Inter-American Development Bank, the Canadian International Development Agency, and others. He is proficient in both quantitative and qualitative methods for data collection and analysis and is knowledgeable of IT systems for both. His passion is keeping abreast of the latest thinking around RBM practices and principles, including Managing for Development Results. He has professional experience in China, Egypt, Ethiopia, Japan, Kenya, the Netherlands, Peru, the Republic of Ireland, Uganda, the UK, the USA, and Zambia, as well as academic experience in India. | kstephenson@universalia.com

Anette Wenderoth graduated from the University of Oldenburgh, Germany with a Ph.D. in Socio-Linguistics and B.A. and M.A. degrees in English Language and Literature, Pedagogy, and Psychology. Following six years of project experience with Universalia in the Balkans, Dr. Wenderoth joined the Ottawa Office where she has completed a series of evaluations with UNIFEM, as well as an evaluation of CIDA's country program in Malawi. She co-edited the 25th Anniversary Universalia publication, *International Monitoring and Evaluation: 25 years of Lessons Learned*. | awenderoth@universalia.com

INDEX

A

Abelard and Heloise, 16

Accountability, 7, 35, 45,
47, 48, 49, 50, 51, 53, 54,
55, 56, 57, 58, 59, 66, 67,
68, 70, 85, 107, 108, 125,
132, 140, 145, 146, 150,
151, 165, 193, 197, 271,
281, 288, 296, 299, 302,
305, 306, 307, 311, 312

act, 132

for results, 29, 46, 47

framework. *See* RMAF

Accra Agreement, 40

Adaptation, 130

Adoption, 36, 173, 221,
227, 233, 253

Adrien, Marie-Hélène, 3

Afghanistan, 31, 325

African and Caribbean
Development Banks,
272

African Evaluation
Association, 302

African Union, 14

Agenda 21, 213

Agroecology, 163, 164, 170

Aid relationship, 23, 34

Algeria, 299

American Evaluation
Association (AEA), 114

Anderson, Gary, 3, 10,
317, 330

Appreciative inquiry, 149

Arab, 298, 299, 300, 302,
303, 305, 306, 312, 313,
314, 329

Arab Bank for
Economic
Development, 300

Nation, 16

Reform Forum, 298

Reform Index, 298

World, 16

- Arab Human
Development Reports
(AHDRs), 298
- Assessment
- Economic impact, 169,
170, 176, 177, 180,
198, 204, 259
 - External, 136
 - Integral Accounting
and Footprint, 230
 - Millennium Ecosystem,
215
 - Organizational, 10, 13,
114, 115, 117, 118,
119, 121, 122, 123,
130, 137, 145, 146,
153, 155, 232, 235,
330, 331, 334
 - Self, 10, 135, 138
- Attribution, 55, 108, 193,
247
- Audit, 50, 67, 150, 227,
237, 309, 320, 324
- Australasian Evaluation
Society (AES), 302
- Australia, 38, 310, 318
- B**
- Bahrain, 298, 306, 309, 310
- Balanced Scorecard, 133,
134, 142, 143, 149, 153
- Beijing Platform for
Action, 271
- Beneficiaries, 7, 48, 71, 72,
105, 124, 151, 181, 189,
194, 200, 286, 290, 311
- Benefits
- intangible, 14, 243, 246,
249, 257, 258
 - tangible, 248
- Best practice, 31, 36, 140
- Biodiversity, 163, 166, 168,
208, 211, 214, 221, 225,
237, 238
- Bonn Convention on
Migratory Species, 211
- Brazil, 39, 209, 213
- Broadband connectivity,
256
- Brundtland Report, 213
- Burke & Litwin Model,
128
- C**
- Canada, 14, 110, 114, 132,
228, 233, 262, 310, 318,
332

- Canadian Evaluation Society (CES), 114
- Capacity
 - analysis, 30
 - checklist, 123
 - development, 182, 195, 196, 302, 331
- Caribbean, 182, 290
- Cartagena Protocol on Biosafety, 214
- Case study, 143, 200, 263
- Case Study
 - case studies, 109, 178, 184, 314
- Causal Model of Organizational Performance and Change, 124, 128
- Cautionary observation, 96
- Ceccarelli, Salvatore, 11, 166, 167, 168
- Center for Effective Philanthropy, 131, 143, 151
- Central authority, 56
- Certification, 198, 320
- Champions, 86
- Chiapas, 209
- Chicoasen Dam, 209
- China, 41, 185, 334
- CIDA, 9, 10, 114, 145, 256, 272, 331, 333, 334
- Climate change, 6, 23, 28, 38, 161, 180, 214, 237, 299
- Climate change, 208, 213, 214
- Club of Rome, 207, 220
- Colombo Plan, 5
- Communication, 29, 85, 95, 98, 105, 152, 239, 243, 251, 260, 274, 331
 - programs, 230
 - theory, 251
- Community
 - Engagement, 6
- Community Development Resource Association (CDRA), 22
- Compliance, 7, 12, 47, 51, 53, 56, 70, 207, 227, 228, 236
- Conflict of interest, 193

Context

analysis, 31, 32

Convention against

Desertification and
Drought, 214

Convention on Biological

Diversity (CDB), 213

Convention on

International Trade in
Endangered Species
(CITES), 211

Cooney, Gerry, 3

Cooperative Group on

International
agricultural Research
(CGIAR), 177

Coordination, 62, 95, 96,

104, 107, 111, 152, 200

Corporate social

responsibility (CSR),
153

Corruption, 307

Costa Rica, 183, 333

Country

context, 30

knowledge, 26, 31

Cuba, 183

D

Darfur, 14

Deep ecology, 217, 223,
235

Demand side, 35

Democratization, 36

Denmark, 233

Department of Economic
and Social Affairs
(OSAGI), 268

Department of Foreign
Affairs, 14

Developing country

less developed, 23, *See*
names of specific
countries

Development

capacity, 4, 6, 7, 16, 19,
20, 23, 25, 26, 27, 28,
29, 32, 33, 34, 35, 36,
37, 38, 39, 40, 41, 43,
113, 182, 195, 196,
307, 333, 334

organizational, 5

sustainable, 213

Development Agencies, 4,
6, 8, 9, 14, 31, 39, 43, 57,
173, 232, 313

- Development Assistance
Committee, 24, 113
- Development
Interventions, 46, 113
- Digital Divide, 252, 254,
255
- Disciplinary
Methodology, 203
- Djibouti, 299
- Document Review, 197,
200
- Donor, 6, 16, 21, 23, 24, 29,
31, 32, 33, 34, 35, 40, 41,
42, 49, 50, 56, 59, 65, 66,
89, 90, 91, 110, 136, 137,
142, 149, 153, 188, 189,
190, 193, 194, 195, 196,
197, 198, 240, 253, 272,
282, 289, 300, 305, 306,
307, 312, 313, 314, 324
- Due Diligence, 137, 228
- E**
- Earth Summit, 213
- ECD, 306, 308, 309, 312,
314
- Ecodevelopment &
Ecoefficiency, 222
- Ecological Footprint, 231,
234
- Ecological Pre-Eminence,
217
- Economic Indicators, 230,
231
- Economic Pre-Eminence,
216
- Effectiveness, 9, 13, 22, 24,
32, 33, 38, 101, 102, 103,
104, 117, 118, 122, 129,
131, 133, 139, 150, 175,
182, 239, 245, 251, 253,
285, 300, 330, 333
- Egypt, 296, 299, 300, 306,
307, 309, 310, 329, 334
- Egyptian Evaluation
Society, 302
- Environment
conference on, 209, 210,
211, 213, 214
crisis, 208, 212, 213
environmental audit,
227
environmental
emergency measures
programs, 229
environmental impact
assessment, 226

- environmental
 - monitoring programmes, 227
- environmental
 - recovery programs, 229
- environmental systems
 - verification, 227
- external, 125, 130, 131
- protection, 161, 211, 228
- Ethics, 134, 153, 154, 225, 307
- Ethnographic Approach, 285, 292
- Europe, 57, 228, 302, 318, 330, 331
 - European Community, 228
- European Evaluation Association (EES), 302
- Evaluation
 - association, 301
 - compliance, 12, 207
 - country-led, 322
 - economic approach to, 171
 - formative, 110
 - horizontal, 191
 - multi-methods, 178
 - policy, 305
 - program, 178
 - questions, 12, 172, 188, 282
 - responsive program evaluation, 178
 - utilization-focused, 172, 184
 - workshop, 185, 191
- Expert opinion, 319
- F**
- Federal Accountability Act, 132
- Feminist
 - approach, 291
 - perspective, 273
- Financial Viability, 129, 139, 142
- Fisher, Ronald Aylmer, 162
- Framework, 7, 10, 13, 46, 47, 53, 64, 76, 103, 104, 105, 106, 110, 120, 125, 127, 128, 129, 130, 131, 134, 135, 138, 139, 149,

- 150, 151, 152, 153, 154,
162, 179, 180, 185, 186,
197, 200, 214, 220, 230,
232, 256, 258, 263, 266,
272, 273, 280, 296, 308,
331
- conceptual, 131, 230
- development results,
64
- foundation
performance
assessment, 131
- livelihoods, 179
- logical, 279
- management results, 64
- MTEF, 310
- social relations, 280
- the Seven-S, 123, 124
- Universalial/ IDRC, 154
- France, 233, 318, 331
- Future Funding, 166, 192,
201
- G**
- Garven, Katherine, 4, 5, 8,
9, 10, 15, 75, 265, 330
- Gates Foundation, 39, 177
- Gender
- analysis, 280
- and development, 274
- conceptualization of,
15, 274, 275, 283
- disaggregated data, 279
- equality, 15, 153, 154,
265, 266, 267, 268,
270, 271, 272, 273,
274, 275, 278, 279,
280, 281, 282, 283,
284, 286, 287, 288,
289, 290, 292, 293,
331, 334
- equity, 268, 269, 271
- mainstreaming, v, 5,
265, 268
- sensitivity, 289
- social construct of, 268
- Genetic Engineering, 163,
164, 169
- Germany, 233, 318, 334
- Global Environmental
Facility, 213
- Global Peace and
Security, 14
- Global Programme of
Action, 214

Globalization, 225, 238,
295

Goals, 9, 65, 78, 79, 80, 82,
83, 84, 88, 92, 103, 109,
110, 116, 118, 159, 168,
178, 182, 193, 200, 226,
262, 283, 290

Gomez, Ricardo, 5, 239,
330

Governance, 22, 23, 53, 77,
78, 79, 94, 107, 128, 197,
215, 282, 296, 297, 299,
302, 306, 307, 308, 312,
314

Grandi, Silvia, 4, 8, 9, 75,
331

Green, Ian, 318

H

Harvard, 10, 118, 133, 267,
280, 318, 324, 330, 333

HIV/AIDS, 63

Horton, Doug, 5

Human rights, 41, 108,
153, 154, 266, 269, 273,
297, 332, 333, 334

Humanitarian
Accountability
Partnership, 150

Hydric Footprint and
Carbon Footprint, 231

I

IAPE, 302

ICT evaluation, 239, 241,
243, 244, 245, 246, 251,
262, 264

IDEAS, 16, 299, 302, 329,
330

IDRC, 10, 77, 114, 123,
129, 130, 131, 139, 154,
167, 254, 256, 273, 288,
297, 331

Incentives, 28, 30, 40, 52,
59, 65, 68, 125, 228, 233,
308, 311

India, 39, 185, 201, 253,
263, 334

Indicator, 56, 69, 73, 140,
141, 142, 253, 269

Indicators

objectively verifiable
indicators, 52

Indonesia, 310, 333

Indonesian, 38

InfoDev, 255, 256, 259

- Information and
Communication
Technologies, 239, 259
- Innovation, 22, 118, 163,
164, 168, 184, 231, 325,
331
- Integrated Natural
Resource Management,
163
- Intelligence
agencies, 31
community, 31
- Inter-American
Development Bank,
181, 333, 334
- International
Development, 3, 4, 5, 7,
9, 24, 25, 31, 39, 41, 45,
50, 55, 66, 75, 76, 99,
125, 155, 241, 252, 265,
266, 273, 274, 319, 322,
331, 332
- International Financial
Institutions, 114
- International Food Policy
Research Institute, 178,
180
- International Program
Evaluation Network,
302
- International Service for
National Agricultural
Research, 182, 192, 331,
332
- International
Telecommunications
Union, 255, 256
- International Union for
the Conservation of
Nature, 211
- IOCE, 302
- Iraq, 299
- IT Value Perception
Model, 247, 248
- Itaipu Dam, 209
- IUCN, 13, 211, 232, 331,
333
- J**
- Jordan, 299, 305, 306
- K**
- Kingdom of Saudi Arabia,
298
- Knowledge, 10, 12, 20, 25,
26, 27, 29, 31, 32, 37, 41,
46, 66, 71, 72, 87, 104,
122, 127, 144, 149, 168,
175, 180, 187, 190, 198,
202, 203, 208, 242, 243,
250, 251, 254, 299, 305,

- 306, 312, 314, 320, 321,
324
- KSA, 298, 299
- Kuwait, 298, 299, 306, 309,
310
- L**
- Latin America, 182, 330,
333
- Leadership, 7, 33, 39, 86,
94, 105, 108, 146, 153,
272
- Learning, 7, 11, 24, 25, 26,
29, 37, 45, 47, 48, 50, 54,
55, 59, 65, 68, 69, 70, 72,
97, 110, 134, 137, 146,
155, 175, 186, 190, 202,
203, 256, 293, 312, 313,
314
- tool to enhance
learning, 202
- Lebanon, 299
- Less developed countries,
26
- Lessons Learned, 54, 65,
83, 201, 257, 293, 318
- Libya, 299
- Logframe, 40, 46, 53, 54,
55, 56, 62, 72, 73
- Logframes, 6, 29, 56, 72
- Lusthaus, Charles, 3, 4, 5,
8, 9, 75, 113, 192, 317,
331, 332
- M**
- M&E, 10, 16, 107, 172, 295,
296, 297, 300, 301, 303,
304, 305, 306, 307, 308,
312, 313, 314, 334
- Malcolm Baldrige Model,
126
- Malmö Declaration, 215
- Man and Biosphere
Programme, 220
- Management
- natural resource, 219,
236
 - process, 127
 - results-based
management, 31, 45,
133, 149, 154
- Management for Results,
45, 51, 56, 65, 73
- Managing for Results,
132, 300, 323
- Market Driven, 89
- Mauritania, 299
- Measures, 10, 57, 131, 133,
134, 135, 140, 142, 143,

- 144, 148, 149, 151, 219,
228, 229, 230, 231, 233,
245, 248, 263, 269, 284,
290, 300, 306
- Mediterranean Action
Plan, 211
- MENA
- Middle East and North
 Africa, 16, 295, 296,
 297, 298, 302
- Mexico, 209, 228, 234, 333
- Micaro, Elisabetta, 5, 265,
332
- Middle East, 5, 16, 298,
299, 302, 330
- Millennium Development
Goals, 39, 41, 254, 266,
299
- Millennium Ecosystem
Assessment, 215
- Mission, 78, 79, 86, 87, 92,
151, 184
- Mobile telephony, 256
- Model
- Burke & Litwin, 128
- Malcolm Baldrige, 126
- Six-Box, 126
- Monitoring, 12, 31, 48, 83,
97, 107, 109, 110, 134,
182, 184, 187, 211, 221,
227, 237, 256, 271, 296,
297, 299, 307, 310, 323,
324, 333, 334
- Montreal protocol on
ozone depleting
substances
- Ozone, 13
- Montreal Protocol on
Substances that
Deplete the Ozone
Layer, 212
- Morgan, Peter, 4, 6, 19,
109, 332, 333
- Morocco, 299, 306, 309,
331
- Most Significant Change,
71
- Motivation, 107, 129
- MS Swaminathan
Research Foundation,
201
- MTEF, 310
- Multilateral agencies, 46
- Multilateral Organization
Performance

- Assessment Network.
See Network
- Multiple data sources,
109, 321
- N**
- Nairobi Declaration, 214
- National Environmental
Policy Act, 226
- Natural Resource
Management, 219
- Network, 8, 75, 87, 88, 96,
105, 113, 149, 152, 187,
190, 260, 313, 319, 320
- M&E, 16, 301
- MOPAN, 133
- NGO, 9, 10, 16, 39, 66, 104,
105, 110, 114, 145, 148,
150, 221, 232, 233, 234,
272, 282, 303, 319, 331
- North Africa, 5, 16, 302
- Not-for-profit
 - agencies, 76
 - sector, 8, 75, 77, 89, 108,
111
 - world, 86, 119, 151
- O**
- Objectively Verifiable
Indicators, 52, 57, 60,
67, 71
- OECD, 22, 24, 30, 113, 114,
256, 300
- Office of Internal
Oversight Services, 47
- Office of the Special
Advisor on Gender
Issues and the
Advancement of
Women, 268, 269
- Oman, 298
- Organization
 - development, 58, 63
 - performance, 133
- Organizational
Development, 5, 21, 23,
122, 125, 185, 331, 333
- Outcome Mapping, 149,
257, 287, 288, 297
- Outcomes, 30, 32, 35, 46,
55, 56, 64, 66, 67, 71, 96,
104, 105, 108, 132, 154,
175, 187, 193, 199, 241,
247, 256, 257, 258, 261,
262, 263, 264, 288, 311

- Outputs, 29, 32, 45, 53, 55,
56, 63, 66, 71, 72, 83, 92,
125, 132, 200, 201, 241,
243, 245, 250, 257, 258,
262, 264
- Ownership, 19, 22, 27, 33,
34, 35, 39, 41, 50, 92,
107, 108, 120, 137, 146,
155, 300, 307, 312
- P**
- Pakistan, 325
- Pakistani, 38
- Palestine, 15, 299
- Panama, 183
- Papa Andina Regional
Partnership Program,
190
- Paradigms, 12, 159, 164,
216, 222, 223, 225, 226,
232, 251
- Paraguay, 209
- Paris Declaration, 39, 65,
133, 299, 300
- Participatory Impact
Pathways Analysis,
186, 187
- Pather, Shaun, 5, 239, 333
- Peace, 14, 109
- Pérez Gil, Ramon, 5
- PETS, 310
- Pilot approach, 310
- Poverty, 28, 37, 40, 41,
117, 177, 178, 179, 180,
259, 268, 283, 297, 298,
317
- Prahbu Pingali, 177
- Problems with
Economics, 170
- Productivity paradox,
244, 245, 252
- Program management, 48,
49, 56, 60, 67, 68, 72
- Public-Private
Partnerships, 152
- Q**
- Qatar, 298
- Qualitative, 10, 57, 58,
140, 143, 175, 178, 180,
243, 284, 285, 320, 321,
334
- Quantitative, 10, 11, 57,
58, 140, 141, 143, 170,
173, 175, 178, 180, 199,
227, 253, 262, 264, 284,
285, 291, 320, 334
- Quesnel, Jean, 319

R

- R.A. Fisher, 11
- Randomised control trials, 162
- Rates of usage, 253
- Ray Rist, 202, 308, 323, 324
- RBAS, 305
- RBM, 4, 7, 8, 10, 40, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 70, 71, 72, 73, 133, 300, 323, 334
- competitive RBM, 55, 59, 61, 62, 66, 68, 71
- false RBM, 51, 52, 53, 55, 56, 59, 67, 70
- improvement-focused RBM, 55, 60, 65, 68
- strategic RBM, 61, 62, 64, 65, 71, 73
- Reagan, Ronald, 4
- Reductionist paradigm, 169
- Reflect and Learn, 123
- Regional Seas Agreement, 211
- Relac, 302
- Relevance, 24, 25, 26, 36, 54, 67, 77, 103, 104, 105, 106, 118, 129, 139, 203, 212, 234, 285, 288, 289
- Resource-Based Theory, 251
- Results Chain, 8, 13, 287
- Rights-Based/Equity Approach, 291, 292
- Rio de Janeiro, 213, 236
- Risk Analysis, 228
- RMAF, 132
- Rockefeller Foundation, 113, 123, 151, 332
- Rojas, Katrina, 4, 5, 113, 265, 333

S

- Scriven, Michael, 317, 318
- Six-Box Model, 125, 126
- Somalia, 299
- South Africa, 22, 215, 263, 333
- Stakeholders, 28, 48, 52, 54, 71, 72, 99, 100, 101, 103, 105, 107, 137, 140, 142, 144, 146, 153, 154, 159, 171, 172, 173, 175, 176, 178, 181, 183, 184,

- 185, 186, 187, 188, 189,
190, 191, 194, 200, 201,
203, 204, 264, 292, 293,
294, 304, 307, 309, 325
- Stephenson, Ken, 4
- Stockholm summit, 210,
213, 227
- Strategic planning, 29, 92
- Sudan, 299
- Support points, 39
- Surveys, 144
- Sustainability, 9, 23, 32,
103, 104, 106, 124, 125,
159, 163, 165, 184, 217
- Sustainable development,
104, 213, 214, 215, 224,
225, 236, 237, 242, 254,
295
- Swiss Agency for
Development and
Cooperation, 190
- Syria, 299, 305
- Systematic, 55, 119, 120,
121, 123, 133, 178, 227
- T**
- Tbilisi, 211
- Technical Assistance, 26,
39, 41, 311, 312, 313, 333
- Terms of reference, 188,
189, 277, 281, 282, 292
- Thatcher, Margaret, 4
- The Netherlands, 38, 331
- Theory
of change, 37, 54, 288,
323
of reasoned action, 251
- Transaction Costs, 9, 96,
97, 98, 104, 152
- Transparency, 52, 194,
197, 299, 302, 312
- Tunisia, 299
- U**
- UK, 4, 125, 162, 233, 332,
334
- UN Conference on the
Human Environment,
209
- UN Security Council, 38
- UNDP, 41, 149, 266, 268,
305, 306, 307, 310, 332,
333
- UNEP, 13, 210, 211, 214
- UNICEF, 149, 297, 299,
331

- United Arab Emirates,
298
- United Nations
Development Fund for
Women, 272
- United Nations
Evaluation Group
UNEG, 266
- United Nations Science,
Culture and Education
Organization's
UNESCO, 210
- Universalia, 3, 4, 5, 7, 9,
10, 13, 14, 15, 16, 19, 43,
58, 75, 114, 129, 130,
131, 139, 145, 149, 154,
155, 232, 265, 266, 267,
271, 272, 282, 284, 287,
289, 292, 293, 294, 307,
317, 319, 326, 329, 330,
331, 332, 334
- UNRWA, 15, 330
- USA, 4, 226, 228, 253, 295,
308, 331, 333, 334
- USAID, 46, 333
- Utilization-Focused
Evaluation, 190
- V**
- Values, 15, 30, 90, 91, 94,
103, 123, 129, 130, 175,
223, 262, 275, 292, 320
- Venezuela, 183
- Vienna Convention for
the Protection of the
Ozone Layer, 212
- W**
- Ways of knowing*, 320, 321
- Wenderoth, Anette, 5,
265, 334
- WHO, 149
- Whole-of-government,
309, 310
- Women and
Development, 274
- Women in Development,
274
- World Bank, 22, 38, 99,
148, 149, 230, 254, 256,
259, 261, 272, 305, 306,
309, 312, 329, 333
- World Big Dams
Congress, 209
- World Conference on
Women, 270

- World Development
Report, 41
- World Summit on
Information Society
(WSIS), 242
- World Summit on
Sustainable
Development, 215
- World
Telecommunication
and ICT Development
report, 259
- Y**
- Yemen, 299, 300
- Yemini, 37
- Z**
- Zimbabwe, 311

