unicef

Evaluation for equitable development results

unicef

Evaluation for equitable development results

ISSUE # 1:   New trends in development evaluation. Published by UNICEF and IPEN, 2006.

ISSUE # 2:   Bridging the gap: The role of monitoring and evaluation in evidence-based policy making. Published by UNICEF in partnership with the World Bank, IDEAS, DevInfo and MICS, 2008.

ISSUE # 3:   Country-led monitoring and evaluation systems. Better evidence, better policies, better development results. Published by UNICEF in partnership with the World Bank, IDEAS, IOCE, UNECE, DevInfo and MICS, 2009.

ISSUE # 4:   Country-led monitoring and evaluation systems. Watch and listen international keynote speakers. DVD published by UNICEF in partnership with IDEAS, IOCE, WFP, OECD/DAC Network on development evaluation and DevInfo, 2009.

ISSUE #5:   From policies to results: Developing capacities for country monitoring and evaluation systems. Published by UNICEF in partnership with DevInfo, IDEAS, ILO, IOCE, The World Bank, UNDP, UNIFEM, WFP, 2010.

ISSUE #6:   How to design and manage Equity-focused evaluations. UNICEF, 2011.

Disclaimer:

This book complements the following material available at **www.mymande.org**
- **Manual on "How to design and manage Equity-focused evaluations"**,
  available free of charge at: http://www.mymande.org/?q=virtual
- **Resource center** available, free of charge, at:
  http://www.mymande.org/?q=equity_focused_evaluations_intro
- **Webinar series** with world-level experts available, free of charge, at:
  http://www.mymande.org/?q=equity_focused_evaluation&x=cl

Evaluation for equitable development results

Editor
Marco Segone

Authors

Michael Bamberger	Karen Kirkhart
Katrina Bledsoe	Donna Mertens
Thania de la Garza	Janice Muir
Soma de Silva	María Fernanda Paredes
Janie Eriksen	Michael Quinn Patton
Oscar Garcia	Shravanti Reddy
Jennifer Greene	Martin Reynolds
Francisco Guzman	Patricia Rogers
Katherine Hay	Jim Rugh
Gonzalo Hernández	Marco Segone
Rodney Hopson	Juha Uitto
Richard Hummelbrunner	Brenda Valdez
Saville Kushner	Bob Williams

Contents

---

1   For additional details on methodological implications, please see Bamberger and
     Segone, 2011, *How to design and manage Equity-focused evaluations* , UNICEF

Preface, Editorial
& Acknowledgements

PREFACE

by the Director of Evaluation, UNICEF

Despite the economic downturn of 2008 and its lingering effects, spectacular headway has been made in reducing global poverty over the past two decades. Progress towards meeting the Millennium Development Goal of halving the proportion of people living in poverty is on track, with the global poverty rate expected to drop below 15% by 2015, far below the target of 23%. Each year, more people escape extreme poverty, and more countries "graduate" to middle or high income status.

Yet despite the rapid progress in reducing poverty, inequalities are increasing both between countries and within countries. A recent UNICEF report on global inequality noted that the richest population quintile enjoys 83 percent of global income, with just a single percentage point going to those in the poorest quintile. As the UN's Secretary-General has written, introducing *The Millennium Development Goals Report 2011*, "Progress tends to bypass those who are lowest on the economic ladder or are otherwise disadvantaged because of their sex, age, disability or ethnicity". A growing body of research confirms that high levels of inequality in the distribution of income, power and resources can slow poverty reduction, exacerbate social exclusion and provoke political and economic instability. Even in rich countries, inequality is dysfunctional, as Richard Wilkinson and Kate Pickett so convincingly demonstrated with the mass of evidence presented in their influential book, *The Spirit Level*.

It is therefore high time to put equity at the center of efforts to promote development. Addressing a high level meeting on the Millennium Development Goals in Tokyo last year, UNICEF's Executive Director, Tony Lake, put the point eloquently. He declared: "There can be no true progress in human development unless its benefits are shared – and to some degree driven – by the most vulnerable among us ... the equity approach is not only right in principle. It is right in practice".

In the same vein, it is an appropriate moment to ask whether evaluation as a discipline and evaluators as a profession are addressing equity issues in ways which are indeed right in principle and right in practice. Some of the answers can be found in the present volume, which brings together a tremendous richness and diversity of evaluation thinking and experience. While a number of the papers included

in the collection touch on approaches and methods already familiar to evaluators, the challenge of addressing the question of equity has helped to demonstrate renewed relevance and establish fresh perspectives. Several essays showcase examples of evaluations addressing equity issues, providing a valuable source of inspiration.

I would therefore like to take this opportunity to express my gratitude to all the contributors to this volume, which I believe is a landmark publication on a topic of central importance. I would especially like to express my appreciation of the fruitful contribution made by my colleague, Marco Segone, in pulling together the present volume, and in organizing the companion series of webinars, which has already provided so much in the way of ideas and illumination for a worldwide community of participants.

In the end, the vital test is whether evaluation can truly help the international community to achieve equitable development results. At the meeting in Tokyo, Mr Lake called for "better monitoring and evaluation of results, to see what is working and where further resources should be focused". Evaluation can indeed help to guide investment towards equitable outcomes. But perhaps even more importantly, the equity approach renews the challenge to evaluators to ensure that the voices of the poorest and most marginal people in society are heard and that their views count in decisions affecting their future. This is perhaps the simplest challenge facing evaluation for equitable development results – but it is probably also the hardest to achieve. I hope the essays in this collection will help evaluators and others rise to meet this test.

Colin Kirk,
Director, UNICEF Evaluation Office

PREFACE

by the President and Vice-President of the International Organization for Cooperation in Evaluation (IOCE)

Achieving greater social equity is increasingly becoming a common goal of governments, civil society organisations and development partners alike. Making development benefits reach the marginalized and the disadvantaged is becoming an integral strategy of national development plans as well as programmes of cooperation of the United Nations Agencies and multilateral and bilateral organizations. Groups of people have been marginalized and disadvantaged for reasons that are historical, cultural and political, among others. These reasons are deep rooted and intricately intertwined with power structures, knowledge levels, belief systems, attitudes and values of societies. They have been barriers to equitable social development. Development programmes and projects have typically tended to be designed with insufficient understanding of these issues. Consequently, development results have often benefited the most advantaged and the better able while only percolating in drops to those who are deprived, thereby perpetuating the inequities.

To penetrate the barriers to social equity, UNICEF is reorienting its programmes of cooperation with governments to achieve greater equity for the most deprived and marginalized children who are a highly vulnerable segment of the population. To achieve successful development results from such programmes, evaluation is as important as are appropriate designs and good implementation. Equity-focused evaluations face certain methodological challenges to address issues related to inequities and their deep rooted causes. In this context, the International Organisation for Cooperation in Evaluation (IOCE) is glad to see the publication of this volume which brings together reflections on the linkages between evaluation and equity as well as diverse methodological approaches to evaluating programmes that promote equity. It brings together a range of methodological approaches of evaluations covering design elements, process elements such as transformative approaches and overall approaches such as systems approaches. It also includes evaluations of a range of interventions such as governance and international development, national development programmes, community based programmes and public policies from widely differing contexts and communities.

It is the mission of IOCE to promote the development of theories and strengthening of evaluation practices that are socially relevant and respect the diversity of cultures, norms and needs. IOCE encourages advancement of theories to address emerging needs such as the approaches to evaluating social equity programmes. Such theoretical advances also enable evaluation associations to better facilitate professional development activities in evaluation that are relevant to the national development contexts.

Soma de Silva   *Jim Rugh*
President, IOCE   *Vice President, IOCE*

PREFACE

by the President of the International Development Evaluation Association (IDEAS)

As President of IDEAS, I am especially pleased to be able to endorse this book and to recommend it to the development community. The issue of equitable development results is increasingly on the radar-screen of policymakers, program managers, and donors. And while the conventional understanding of equitable development is asso-ciated with the Millennium Development Goals (MDGs), there are broader implications that are addressed in this volume. Equitable development is not just working to achieve the eight goals of the Millennium Declaration in 2015, but conceptually and politically it carries us into the arenas of structural inequity, regional inequity, and the inequity that emerges from unfair trade practices, unfair currency manipulation, authoritarian regimes, and non-democratic governance. This book opens up the conversation to such issues and for this it is to be commended.

I wish to thank UNICEF for their initiative and perseverance in bringing this book to fruition. It is a contribution to the evaluation community and for this, we all own UNICEF a strong thank you.

*Ray C. Rist
President, IDEAS*

EDITORIAL

This publication aims to stimulate the international debate on how the evaluation function can contribute to achieving equitable development results by conceptualizing, designing, implementing and using evaluations focused on human rights and equity.

It does so by offering a number of strong contributions from senior officers in institutions dealing with international development and evaluation. These are: UNICEF, UNDP, UNWomen, ILO, IDRC, the International Development Evaluation Association (IDEAS) and the International Organisation for Cooperation in Evaluation (IOCE); as well as senior Government representatives responsible for evaluation systems in their country, such as CONEVAL in Mexico.

The book is divided into three parts. Part I presents the relationship between evaluation and human rights and equity. Part II focuses on the methodological implications in design, implement and use of Equity-focused evaluations; and part III presents few examples of Equity-focused evaluations.

In Part I, **Segone** introduces Equity-focused evaluations by explaining what equity is and why equity matters. He continues by defining Equity-focused evaluations, what's the purpose and positioning equity-focused evaluations as a pro-equity intervention. He argues that, while Equity-focused evaluations pose new challenges, they also constitute an opportunity to make evaluation an action for change to achieve development results with an equity focus.

Reddy, Eriksen and **Muir** explain why it is important to integrated human rights and gender equality in evaluation, and the implications in doing so at each of stage of the evaluation process. They also present the United Nations Evaluation Group (UNEG) – a professional network that brings together the units responsible for evaluation in the UN system – response to a gap in the integration of these important dimensions in evaluation of the UN's work: a handbook containing practical guidance for evaluators.

Mertens addresses the challenges of planning, implementing and using evaluations that emerge when human rights is the starting point for policymakers, funders, programme developers and evaluators. She explains that situating oneself as an evaluator in a human rights position requires re-thinking how evaluation is conceptualized, practiced, and used in international development. Finally, she makes the argument that re-framing of evaluation from this starting

point, based on human rights, provides fertile ground for obtaining meaningful answers to questions about the efficacy of international development interventions.

Hay examines how principles drawn from feminist and other research traditions and theories, can be used in practice to inform the understanding of programme theory; shape evaluation design and methods; negotiate judgment of success; guide practice; and, guide choices and opportunities for influence. She suggests that principles generated from feminist theory can help evaluation to play a stronger role in understanding how societies change and which policies and programmes show promise in shifting norms and achieving equity.

Hopson, Kirkhart and Bledsoe suggest that the good intentions of Equity-focused evaluations must be tempered by cautions. This concern flows from a legacy of research and evaluation that has exerted colonizing influences over Indigenous and minoritized populations. The opening section covers the context of development, evaluation, and culture. The second section argues that efforts to decolonize evaluation must begin with epistemology. A third section examines the implications of decolonization for evaluation method. Within the paper, a scenario is provided based on a development project in southern Africa. The authors close with implications and cautions for how evaluation generally, and more specifically, Equity-focused evaluations may perpetuate colonizing assumptions and aims.

In part II, **Bamberger** discusses strategies and methods for evaluating how well development interventions address and achieve equity objectives. He starts by highlighting the importance of distinguishing between *simple* equity-focused projects, and *complex* equity-focused policies and other national level interventions as these affect Equity-focused evaluation design options. He then presents different approaches and tools to evaluate equity-focused impact at policy level and at project level, highlighting the importance of mixed-methods.

Patton explains that Developmental evaluation supports innovative intervention *development* to guide adaptation to emergent and dynamic realities in complex environments. He suggests that evaluation for equity and the fostering of human rights, as part of achieving meaningful development results, often occurs in complex adaptive systems. In such situations, informed by systems thinking and sensitivity to complex nonlinear dynamics, Developmental evaluation supports increased effectiveness of interventions, social

innovation, adaptive management, and ongoing learning.

Reynolds and **Williams** argue that Equity-focused evaluations should be instrumental in redressing prevailing inequities of resource-access. In this context, they explain that it is often difficult to appreciate the wider picture of issues relating to resource access, including different perspectives on inequities from different stakeholders. A systems approach to Equity-focused evaluation prompts, firstly, a greater awareness of the interrelated issues of inequities; secondly, an appreciation of different perspectives on inequities, and; thirdly, a reflection on boundaries used to circumscribe our awareness and appreciation of inequities.

Rogers and **Hummelbrunner** discuss some of the key features of equity-focused programmes that programme theory needs to address – in particular, the need to support poor and marginalized people to be agents of their own development, and to address complicated and complex aspects of programmes. They then present the implications of these for developing, representing and using programme theory, arguing that programme theory needs to acknowledge the other factors needed to produce intended outcomes and impacts; support appropriate translation of effective interventions to other contexts, by distinguishing between theories of change and theories of action; highlight differential effects of interventions, and in particular the distribution of benefits; and, support adaptive management of emergent programmes.

Kushner contributes an approach to evaluation that makes it an equitable process. Its focus is methodological, and he suggests that the methodological solution of Equity-focused evaluation is case study. He argues that case study gives us a more systemic/dynamic view of policy and public value, as describing, analysing and understanding the implications of policy-*shaping* is a key task for the case study evaluator who needs to understand the sources and consequences of unequal access to information and power asymmetries in setting the criteria against which interventions will be judged.

Greene presents the character of, and rationale for, in explicit naming and claiming of values in evaluation, including in Equity-focused evaluation. Greene argues that values are present in virtually all aspects of evaluation. The term "values-engagement" is intended to signal explicit attention to values as part of the evaluation process and to the central role that values play in our evaluation practice. From the framing of evaluation questions to the development of an evalua-

tion design and methods, and from the interactions of stakeholders in the evaluation process to the especially important task of making judgments of programme quality, values are a central feature in this approach. Engagement thereby suggests a kind of quiet insistence that questions of value be addressed throughout the evaluation, at every turn and every decision point – so values become interlaced with, knitted and knotted within evaluative thinking and judging.

In part III, **Uitto** and **Garcia** attempt to extract lessons from evaluative evidence gathered from the Assessment of Development Results conducted by the UNDP Evaluation Office in China and Brazil. They focus on the role played by international cooperation, particularly by UNDP and other international partners, in support of equity-focused public policies. After providing a brief overview of historical trends in inequity in Brazil and China, Uitto and Garcia outline the main findings of the evaluations regarding UNDP contributions towards policies that address inequities in the two countries. Finally, they end with a brief section on lessons learned and conclusions.

Guzman summarizes the challenges and lessons learned in including the transformative paradigm in the methodology used for high-level evaluation of the International Labour Organization's (ILO) discrimination strategy. In line with the Human rights and Gender equality approach to offer diverse perspectives to the evaluation, and to promote participation of different groups of stakeholders, the evaluation required setting-up an appropriate mix of qualitative and quantitative methods to gather and analyse data. The evaluation used a mixed-method approach including (but not limited to) desk reviews, interviews, focus groups, surveys, etc. In addition, the evaluation team took into account not only the policy and normative framework but also carefully discerned power relationships, and identified the structural causes of discrimination in employment and occupation.

Hernández Licona, de la Garza, Paredes and Valdez explain that Mexico is a country with prevailing challenges in various dimensions with regard to social inequities. One of the harshest manifestations of the social gaps that persist in Mexico is the lack of opportunities for the indigenous population, which leads to serious limitations for the exercise of their rights and provides evidence of the social inequities that prevail among the population. In their article, the authors present an assessment of the adequacy and results of social policies in order to analyze the situation of the indigenous people and the government response through public policies.

This book is part of a continuous effort led by UNICEF, in partnership with several key stakeholders, to stimulate the debate on how evaluation can contribute to equitable development results, as well as the sharpening of methodological approaches to ensure that interventions designed to enhance equity can be evaluated in a meaningful manner. This book complements the manual "How to design and manage Equity-focused evaluations" published by UNICEF in 2011; the electronic resource centre managed by UNICEF and UNWomen, available at www.mymande.org; and, the series of webinars with international keynote speakers.

We hope this stream of work will enhance the capacity of the evaluation community to strengthen the relevance and quality of evaluations so as to better inform equitable interventions. I wish you an interesting and inspiring read.

Marco Segone
Editor

ACKNOWLEDGEMENT

Part 1
Evaluation and equity

EVALUATION TO ACCELERATE PROGRESS TOWARDS EQUITY, SOCIAL JUSTICE AND HUMAN RIGHTS[1]

*Marco Segone, UNICEF Evaluation Office;
Co-chair, UNEG Taskforce on National Evaluation Capacities;
and former IOCE Vice President*

What is equity and why does it matter?

The challenge of achieving equitable development results for children

When world leaders adopted the Millennium Declaration in 2000, they produced an unprecedented international compact, a historic pledge to create a more peaceful, tolerant and equitable world in which the special needs of children, women and those who are worst-off can be met. The Millennium Development Goals (MDGs) are a practical manifestation of the Declaration's aspiration to reduce inequity in human development among nations and peoples by 2015. The past decade has witnessed considerable progress towards the goals of reducing poverty and hunger, combating disease and child mortality, promoting gender equality, expanding education, ensuring safe drinking water and basic sanitation, and building a global partnership for development. But with the MDG deadline only a few years away, it is becoming ever clearer that reaching the poorest and most marginalized communities within countries is pivotal to the realization of the goals (UNICEF, 2010c).

Since 1990, significant progress has been made on several MDGs. However, the gains made in realizing the MDGs are largely based on improvements in national averages. A growing concern is that progress based on national averages can conceal broad and even widening disparities in poverty and child development among regions and within countries. In child survival and most other measures of progress towards the MDGs, sub-Saharan Africa, South Asia and the least developed countries have fallen far behind other developing regions and industrialized countries. Within many countries, falling national averages for child mortality conceal widening inequi-

---

1   Based upon Bamberger M. and Segone M. (2011). *How to design and manage Equity-focused Evaluation*, UNICEF.

ties. The same is true for several other indicators, including early childhood development, education, HIV/AIDS and child protection (UNICEF, 2010d). Disparities hamper development not only in low income countries, but also in middle income countries. A UNICEF study conducted in Brazil (UNICEF Brazil, 2003) showed that compared to rich children, poor children were 21 times more likely to be illiterate. But poverty is not the only cause of inequity. According to the same study, compared with white children, black children were twice as likely not to attend school, and children with disabilities were four times more likely to be illiterate compared to children without disabilities.

These marked disparities in child survival, development and protection point to a simple truth. The MDGs and other international commitments to children can only be fully realized, both to the letter and in the spirit of the Millennium Declaration, through greater emphasis on equity among and within regions and countries (UNICEF, 2010c).

What is equity?

For UNICEF "equity means that all children have an opportunity to survive, develop, and reach their full potential, without discrimination, bias or favoritism" (UNICEF, 2010a). This interpretation is consistent with the Convention on the Rights of the Child (CRC), which guarantees the fundamental rights of every child, regardless of gender, race, religious beliefs, income, physical attributes, geographical location, or other status.

This means that pro-equity interventions should prioritize worst-off groups[2] with the aim of achieving universal rights for all children. This could be done through interventions addressing the causes of inequity and aimed at improving the well-being of all children, focusing especially on accelerating the rate of progress in improving the well-being of the worst-off children.

Equity is distinguished from equality. The aim of equity-focused policies is not to eliminate all differences so that everyone has the same level of income, health, and education. Rather, the goal is to eliminate the unfair and avoidable circumstances that deprive children of their rights. Therefore, inequities generally arise when certain population groups are unfairly deprived of basic resources that are available to other groups. A disparity is 'unfair' or 'unjust' when

---

2   As different countries and different organizations use different terminology such as excluded, disadvantaged, marginalized or vulnerable populations, here the term "worst-off groups" is used to refer to those population groups suffering the most due to inequity.

its cause is due to the social context, rather than the biological factors.

While the concept of equity is universal, the causes and consequences of inequity vary across cultures, countries, and communities. Inequity is rooted in a complex range of political, social, and economic factors.

An equity-focused intervention must therefore begin with an analysis of the context in which inequity operates. This analysis informs the design of programme and interventions that are tailored to address the local causes and consequences of inequity. These initiatives must be developed in collaboration with national partners who can help identify culturally appropriate strategies for promoting equity.

Why does equity matter?

Achieving equitable development results...

As explained above, UNICEF states that the MDGs and other international commitments to children can only be fully realized through greater emphasis on equity among and within regions and countries, for the following reasons (UNICEF, 2010c). Firstly, several key international goals for children require universality. One of the most prominent is MDG 2, which seeks universal access to primary education. Logically, this objective can only be met if the children currently excluded, who are the poorest and the most marginalized, are brought into the school system. Similarly, it will be impossible for global campaigns seeking the eradication of polio, or virtual elimination of measles and maternal and neonatal tetanus, to succeed without addressing the poorest communities within countries. Secondly, having reduced the global under-five mortality rate by one third since 1990, countries now have few years to do so again to meet the conditions of MDG 4. Since most child deaths occur in the most deprived communities and households within developing countries, achieving this goal is only possible by extending to them the fight against childhood illness and under-nutrition. Thirdly, breaking the cycle of poverty, discrimination, educational disadvantage and violence experienced by many girls and young women is only possible through equity-focused approaches that eliminate gender-based barriers to essential services, protection and girls' knowledge of their rights. Fourthly, new technologies and interventions can contribute to faster gains for the poor if applied equitably and at scale.

...for socially fair, politically stable and economically strong societies

In *The Spirit Level*, Picket and Wilkinson (2009) show that in richer countries inequity is associated with a wide range of social problems including: levels of trust; mental illnesses; life expectancy; infant mortality; obesity; educational performance; drug use; teenage births; homicides; and, imprisonment rates. In most cases these indicators are not closely related to the per capita income or rate of growth of a country, and so higher rates of growth tend not to be associated with reducing social problems. Also, available evidence for both developed and developing countries does not suggest that inequity is reduced over time by high rates of economic growth. In addition, equity is important for the following reasons (Segone, 2003):

- **Inequity constitutes a violation of human rights**. Inequity remains among the most important human rights challenges facing the world community. A human rights-based approach means that, in the light of the principle of universality and non-discrimination, all children, from birth to childhood and adolescence, boys and girls, of whatever color, race, language or religion and wherever they may live, need to be considered (Santos Pais, 1999). It means that the situation of poor people is viewed not only in terms of welfare outcomes but also in terms of the obligation to prevent and respond to human rights violations. The High Commissioner for Human Rights stated that human rights are about ensuring dignity, equity and security for all human beings everywhere. (UN NGLS, 2002).

- **Inequity is one of the major obstacles in taking advantage of the richness of diversity**. If human beings do not all have the same opportunity, some groups are discriminated against and excluded from society. Inequity means that society is not giving these individuals and groups equal opportunity to contribute to the development of the country. It means that it is focusing mainly on one "cultural model" and is not taking advantage of diverse "cultural models", which can foster societal innovation and creativity.

- **Equity has a significant positive impact in reducing monetary poverty**. Monetary poverty is very sensitive to distribution changes, and small changes in income distribution can have a large effect on poverty. For a given level of average income, education, land ownership etc., an increase in monetary

inequity will almost always imply higher levels of both absolute and relative deprivation and vice versa (Maxwell and Hanmer, 1999).

- **Equity has a positive impact in the construction of a democratic society**. Equity facilitates citizen participation in political and civil life. A citizen's capacity to participate in political and civil life and to influence public policies is linked to his/her income and education. In a political system based on citizen's income, significant income inequity means significant inequity in the political system. This leads to higher inequity in the educational system, due to lower investment in quality education. This means poor children attend lower quality schools and therefore a wider gap is created between education and capacity (the "human capital") acquired by the poor children attending low quality public schools, and the rich children attending high quality private schools. This vicious cycle closes with the inequity in education impacting negatively on income inequity, as income is directly linked to the level of education.

- **Prolonged inequity may lead to the "naturalization" of inequity**. In several countries institutional and historical origins of inequity are multiple, but its persistence, or worsening, over the decades makes inequity something "natural". When inequity is perceived as a natural phenomenon (the so called "naturalization of inequity"), societies develop theoretical, political and ideological resistances to identifying and fighting inequity as a priority. Along the same lines, inequity may even create self-fulfilling expectation and acceptance of lower growth. If workers are paid according to social class, gender or race/ethnicity, rather than by what they achieve, this reduces the incentive to work/earn more.

- **Inequity may lead to political conflict and instability**. Last but not least, unequal opportunities for social groups in society – and perhaps more importantly, inequities as perceived by these groups – are often also a significant factor behind social unrest. This may lead to crime or even violent conflict, as well as lower investment and more waste of resources from bargaining over short-term distribution of rents. Highly polarised societies are unlikely to pursue policies that have long-term benefits for all, since each social group will be reluctant to make long-term commitment, dedicated as they are to secure their own wealth. Along the same line of argument, this instability also reduces

government's ability to react to shocks. The economic costs of external shocks are magnified by the distributional conflicts they trigger, and this diminishes the productivity with which a society's resources are utilised. This is largely because social polarisation makes it more difficult to build consensus about policy changes in response to crisis.

What are the implications for the evaluation function?

The renewed focus on equity poses important challenges – and opportunities – to the evaluation function: What are the methodological implications in designing, conducting, managing and using Equity-focused evaluations? What are the questions an Equity-focused evaluation should address? What are the potential challenges in managing Equity-focused evaluations? This document, together with the manual on "How to design and manage Equity-focused Evaluations", the webinars and electronic resources available at www.mymande.org, represents a first attempt to address these challenges.

Defining Equity-focused evaluations

What is an Equity-focused evaluation?

An Equity-focused evaluation is a judgment made of the relevance, effectiveness, efficiency, impact and sustainability – and, in humanitarian settings, coverage, connectedness and coherence – of policies, programmes and projects concerned with achieving equitable development results. It involves a rigorous, systematic and objective process in the design, analysis and interpretation of information in order to answer specific questions, including those of concern to worst-off groups. It provides assessments of what works and what does not work to reduce inequity, and it highlights intended and unintended results for worst-off groups as well as the gap between best-off and worst-off groups. It provides strategic lessons to guide decision-makers and to inform stakeholders. Equity-focused evaluations provide evidence-based information that is credible, reliable and useful, enabling the timely incorporation of findings, recommendations and lessons into the decision-making process.

Why are Equity-focused evaluations needed?

Equity-focused evaluations look explicitly at the equity dimensions of interventions, going beyond conventional quantitative data to the anal-

ysis of behavioral change, complex social processes and attitudes, and collecting information on difficult-to-reach socially marginalized groups. In addition, Equity-focused evaluations constitute a pro-equity intervention by empowering worst-off groups, as described below.

It is however important to highlight that while some new analytical tools are introduced (particularly the bottleneck supply and demand framework, and systems approaches to evaluate equity interventions in complex environments), most of the Equity-focused evaluation approaches, as well as data collection and analysis techniques, are built on methods which are already familiar to many practitioners in development evaluation. So the emphasis is on refining and refocusing existing approaches and techniques – and enhancing national capacities to use them – rather than starting with completely new ones.

Purposes of Equity-focused evaluations

Equity-focused evaluation contributes to good governance of equity-focused policies, programmes and projects for the purposes explained below. These will vary according to context, interventions content and partner interests, among other factors.

Accountability. Equity-focused evaluation ensures that reporting on relevance, impact, effectiveness, efficiency and sustainability of pro-equity interventions is evidence-based.

Organizational learning and improvement. Knowledge generated through Equity-focused evaluations provides critical input into major decisions to be taken to improve equity-focused interventions.

Evidence-based policy advocacy. Knowledge generated through an Equity-focused evaluation provides evidence to influence major policy decisions to ensure that existing and future policies will enhance equity and improve the well-being of worst-off groups. Equity-focused evaluation provides information that has the potential to leverage major partner resources – and political commitment – for pro-equity programmes/policies.

Contribute to Knowledge Management. Understanding what works and what does not work in pro-equity interventions and ensuring that lessons learned are disseminated to national and global knowledge networks helps accelerate learning, avoid error and improve efficiency and effectiveness. It is important to harvest the evidence-base, particularly resulting from innovative program-

ming to foster equity, to demonstrate what works in diverse country contexts.

Empowerment of worst-off groups. If Equity-focused evaluation is to be truly relevant to interventions whose objective is to improve the well-being of worst-off groups, the Equity-focused evaluation processes must be used to foster wider participation of worst-off groups, facilitate dialogue between policymakers and representatives of worst-off groups, build consensus, and create "buy-in" to recommendations. In addition, involving these groups in Equity-focused evaluation can be empowering. It imparts skills, information and self-confidence and so enhances the "evaluative thinking". It can also strengthen the capacity of worst-off groups to be effective evidence-based advocates. Employing Equity-focused evaluation as a programming strategy to achieve empowerment can be very effective, and it can reinforce the other purposes of evaluation.

National Capacity development for equity-focused M&E systems. Countries (central and local authorities, governmental and civil society organizations) should own and lead their own national equity-focused M&E systems. International organizations should support national equity-focused monitoring and evaluation capacity development to ensure that it is sustainable and that the information and data produced are relevant to local contexts, while being in compliance with M&E standards (Segone, 2009 and 2010).

Equity-focused evaluations as pro-equity interventions

As already seen above, Equity-focused evaluation processes should be used to empower worst-off groups to the maximum extent possible, as well as to ensure that evaluation questions are relevant to the situation of these groups. This has two major implications.

First, equity-focused evaluation should be culturally sensitive and pay high attention to ethics. Evaluators should be sensitive to local beliefs, manners and customs and act with integrity and honesty in their relationships with all stakeholders, including worst-off groups, as stated in the standards for evaluation in the UN System (UNEG 2005). In line with the UN Universal Declaration of Human Rights and other human rights conventions, evaluators undertaking Equity-focused evaluation should operate in accordance with international values. Evaluators should be aware of differences in culture; local customs; religious beliefs and practices; personal interaction

and gender roles; disability; age and ethnicity; and, be mindful of the potential implications of these differences when planning, carrying out and reporting on evaluations. In addition, the evaluators should ensure that their contacts with individuals are characterized by respect. Evaluators should avoid offending the dignity and self-respect of those persons with whom they come in contact in the course of the evaluation. Knowing that evaluation might often negatively affect the interests of some stakeholders, the evaluators should conduct the evaluation and communicate its purpose and results in a way that clearly respects the dignity and self-worth of the worst-off groups.

Secondly, equity-focused evaluation should use participatory and/or empowerment evaluation processes to ensure worst-off groups are involved and/or co-leading the Equity-focused evaluation process starting at the design phase. Participatory Equity-focused evaluation processes should pay particular attention to existing imbalances in power relationships between worst-off groups and other groups in society. This is to avoid worst-off groups participating in the Equity-focused evaluation being merely "providers" of information or even of being manipulated or excluded. Selection of stakeholders in Equity-focused evaluation processes should ensure that the processes and methods used serve to correct, not reinforce, patterns of inequity and exclusion. In addition, Equity-focused evaluations must also be aware of power relations within worst-off groups. In many ethnic minorities and disadvantaged groups, certain sectors are further marginalized on the basis of factors such as age, gender, land ownership, relative wealth or region of origin. Great cultural sensitivity is required to respect cultural norms while ensuring that marginalized groups are able to participate and have access to services.

Equity-focused evaluations should also involve young people and children as appropriate, since young people and children are often among the worst-off groups. The Convention on the Right of the Child provides clear initial guidance for the participation of children in evaluation, when it states that the views of children must be considered and taken into account in all matters that affect them. They should not be used merely as data providers or subjects of investigation (CRC, 1990). Article 13 of the CRC states that children have the right to freedom of expression, which includes seeking, receiving and giving information and ideas through speaking, writing or in print, through art or any other media of the child's choice. Their participation is not a mere formality; children must be fully informed and

must understand the consequences and impact of expressing their opinions. The corollary is that children are free not to participate, and should not be pressured. Participation is a right, not an obligation.

Conclusion

While Equity-focused evaluations pose new challenges, they also constitute an opportunity to make evaluation an action for change to achieve development results with an equity focus. The resources and material UNICEF and partners are making available constitute an important contribution to enhance national capacities in equity-focused evaluations.

References

Bamberger, M., Rugh, J. & Mabry, L. (2006). *Real World Evaluation*. Sage Publications.

Convention on the Rights of the Child (CRC) (1990).

DFID (2008). *Reducing poverty by tacking social exclusion: A DFID policy paper.*

Hart, R. (1992). *Children's Participation: from Tokenism to Citizenship. Innocenti Essays No. 4.* New York: UNICEF.

Henriques, R. (2001). *Desigualdade Racial no Brasil: evolução das condições de vida na* década de 90. Texto para discussão nº. 807. Rio de Janeiro: IPEA.

Hills, J.,Le Grand, J. & Piachaud, D. (Eds) (2001). *Understanding social exclusion.* Oxford University Press.

Kahn, S. (2009). *Topic Guide on Social Exclusion.* Governance and Social Development Resource Centre, International Development Centre. University of Birmingham.

Naudeau, S. et al. (2011). *Investing in young children: An early childhood development guide for policy dialogue and project preparation.* The World Bank.

NGLS (2002). *Go between*, 92. Geneva, Switzerland.

Paes de Barros, R., Henriques, R., & Mendonça R. (2002). *Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil.* Texto para discussão nº. 857. Rio de Janiero: IPEA.

Picciotto, R. (2002). *International trends and development evaluation: the need for ideas.* American Journal of Evaluation. 24 (2): 227-34

Picket, R. & Wilkinson, K. (2009). *The Spirit Level: Why equality is better for everyone.* Penguin Books

Rajani, R. (1999). *Promoting Strategic Adolescent Participation: a discussion paper for UNICEF.* New York: UNICEF.

Rieper, O Leeuw, F & Ling, T (2010). *The evidence book: concepts, generation and use of evidence.* Transaction Publishers.

Salmen, L. (1987). *Listen to the People: Evaluation of Development Projects.* New York. Oxfor University Press.

Santos, P. (1999). *A human rights conceptual framework for UNICEF. Innocenti essay nº. 9.* Florencia, Italia: UNICEF Innocent Research Centre.

Segone, M. (1998). *Democratic Evaluation: A proposal for strengthening the evaluation function in International Development Organizations.* UNICEF TACRO.

Segone, M. (2001). *Child poverty in Niger. Longitudinal, gender, spatial and wealth analysis based on MICS and DHS data,* UNICEF Niger.

Segone, M. (2003). *Equity and diversity in Brazilian childhood and adolescence.* UNICEF Brazil.

Segone, M. (2004). *How to achieve the Millennium Development Goals? Reducing inequity through the celebration of diversity,* UNICEF Global Policy Section.

Segone , M. (Ed.), (2006). *New Trends in Development Evaluation.* UNICEF Regional Office for CEE/CIS and IPEN. Available at http://www.mymande.org/?q=virtual

Segone, M. (Ed.), (2008). *Bridging the gap: The role of monitoring and evaluation in evidence-based policy making.* UNICEF, DevInfo, IDEAS, MICS, The World Bank. Available at http://www.mymande.org/?q=virtual

Segone, M. (Ed), (2009). *Country-led M&E systems.* UNICEF, UNDP, WFP, UNWomen, IDEAS, IOCE. Available at http://www.mymande.org/?q=content/policies-results

Segone, M. (Ed), (2010). *From policies to results: Developing capacities for country monitoring and evaluation systems.* UNICEF, DevInfo, IDEAS, ILO, IOCE, The World Bank, UNDP, UNIFEM, WFP. Available at http://www.mymande.org/?q=content/policies-results

Segone, M. & Bamberger, M. (2011). *How to design and manage Equity-focused evaluations.* UNICEF. Available at http://www.mymande.org/?q=content/policies-results

UNDP. Human development Report (2010). T*he real wealth of nations: Pathways to Human Development.*

UNEG (2011). *Integrating human rights and gender equality in evaluation.*

UNICEF (2003). Report on the Situation of Children and Adolescents in Brazil.

UNICEF (2010a). *Re-focusing on Equity: Questions and Answers.*

UNICEF (2010b). *Seven contributions of the evaluation function to the equity approach.*

UNICEF (2010c). *Narrowing the gaps to meet the goals.*

UNICEF (2010d). *Progress for Children: Achieving the MDGs with Equity.*

UNICEF (2011a). *Equity and reaching the most marginalized: Selected innovations and lessons from UNICEF programmes.*

UNICEF (2011b). Guidance Note. *Equity Marker: Resource tracking for equity-focused programme ming. Division of Policy and Practice & Programme Division. New York.*

United Nations News Centre (2001). *UN and Oxford University unveil new index to measure poverty*, http://www.un.org/apps/news/story.

HUMAN RIGHTS AND GENDER EQUALITY IN EVALUATION[1]

Shravanti Reddy, Evaluation Office, UN Women;
Janie Eriksen, Evaluation Office, UNICEF, and Janice Muir,
Evaluation Office, OIOS; And Members of the UNEG task force on
Human rights & gender equality

Human rights and gender equality (HR & GE) are the fundamental dimensions that guide the work of the United Nations (UN). However, in 2007, the United Nations Evaluation Group (UNEG), a professional network that brings together the units responsible for evaluation in the UN system, recognized a gap in the integration of these important dimensions in evaluations of the UN's work. In response, it set up an HR & GE Taskforce to develop guidance on this issue. This year, the Taskforce has released a handbook containing practical guidance for evaluators that it hopes will serve as a step towards addressing this gap. The handbook is entitled: Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance.[2]

What are human rights and gender equality?

Human rights (HR) are the civil, cultural, economic, political and social rights inherent to all human beings, whatever their nationality; place of residence; sex; national or ethnic origin; colour; religion; language; or any other status. All human beings are entitled to these rights without discrimination. They are universal; inalienable; interdependent; indivisible; equal and non-discriminatory; and expressed in, and guaranteed by, normative frameworks and laws that lay down the obligations of States to act in order to respect, protect and fulfil the human rights and fundamental freedoms of individuals or groups. The term 'duty-bearers' reflects the obligations of the State towards 'rights-holders' which represent all individuals in the concerned State.[3] In addition to the Universal Declaration of Human

---

1   This article is based on: *Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance*, UNEG, 2011

2   The handbook can be found at the following site: http://www.unevaluation.org/HRGE_Guidance.

3   Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance, p 12

Rights (UDHR) – the overarching document that formally recognizes universally agreed human rights – there are nine core international human rights treaties that further delineate and codify the rights contained in the UDHR[4]. The strategy for implementing human rights in UN programming is called the Human Rights-Based Approach to programming.

Gender equality (GE) refers to the equal rights, responsibilities and opportunities of women and men, girls and boys. It 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 concerns, and should fully engage, men as well as women. Equality between women and men, girls and boys is seen both as a human rights' issue and as a precondition for, and indicator of, sustainable people-centred development. It is also an

> ### *UN common understanding of the Human Rights-Based Approach to Programming*
>
> 1. All programmes of development cooperation, policies and technical assistance should further the realization of human rights as laid down in the Universal Declaration of Human Rights and other international human rights instruments.
>
> 2. Human rights standards contained in, and principles derived from, the Universal Declaration of Human Rights and other international human rights instruments guide all development cooperation and programming in all sectors and in all phases of the programme process.
>
> 3. Development cooperation contributes to the development of the capacities of "duty-bearers" to meet their obligations and/or of "rights-holders" to claim their rights.
>
> *Source: http://www.ohchr.org/Documents/Publications/FAQen.pdf*

essential component for the realization of all human rights. Progress toward gender equality requires changes within the family, culture, politics and the economy, in addition to changes in laws and their application. The 1979 Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), the Beijing Platform for Action (BPA) and the Millennium Declaration commit the UN to

---

4   See http://www2.ohchr.org/english/law/index.htm#core

promoting GE in its work.[5] Gender mainstreaming is the strategy adopted by the UN for integrating gender equality in programming.

> ### *1997 UN ECOSOC Resolution on Gender Mainstreaming*
>
> Gender mainstreaming is defined as "the process of assessing the implications for women and men of any planned action, including legislation, policies or programmes, 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 programmes in all political, economic and societal spheres, so that women and men benefit equally and inequality is not perpetuated. The ultimate goal is gender equality."
>
> *Source: http://www.un.org/womenwatch/daw/csw/GMS.PDF, Chap IV A Para 4 I A.*

Why is it important to integrate human rights and gender equality in evaluation?

Evaluations play a crucial role in examining to what extent UN interventions benefit rights-holders (particularly those most likely to have their rights violated), and how they strengthen the capacity of duty-bearers or other actors to fulfil obligations and responsibilities, and strengthen accountability mechanisms and monitor and advocate for compliance with international standards on HR & GE. Evaluation can also shed light on how these processes occur and can call attention to the exclusion of certain groups.

An evaluation that neglects or omits consideration of HR & GE deprives the UN system of evidence about who does (and does not) benefit from its interventions, risks perpetuating discriminatory structures and practices where interventions do not follow UN policy in these areas. It may therefore miss opportunities for demonstrating how effective interventions are carried out.

What does it mean to integrate human rights and gender equality in evaluation?

An evaluation that is HR & GE responsive addresses the programming principles required by a human rights based approach and

---

5   See http://www.un.org/womenwatch/osagi/conceptsandefinitions.htm) and UN (2008), Report on indicators for promoting and monitoring the implementation of human rights, pp. 4-10, paras 5 and 12, http://www2.ohchr.org/english/bodies/icm-mc/docs/HRI.MC.2008.3EN.pdf

gender mainstreaming strategy. It contributes to the social and economic change process that is at the heart of most development programming by identifying and analysing the inequalities, discriminatory practices and unjust power relations that are central to development problems. HR & GE responsive evaluations can lead to more effective interventions and better, more sustainable results.

Given the mandate to support and incorporate HR & GE in all UN work, these dimensions need to be paid special attention when evaluating UN interventions. This requires attention to the interrelated principles of inclusion, participation and fair power relations.

- **Inclusion.** Evaluating HR & GE requires attention to be paid to the beneficiary groups in the intervention under review. Some groups may be negatively affected by an intervention. An evaluation must acknowledge who these stakeholders are, how they are affected, and how to minimize these negative effects.

- **Participation.** Evaluating HR & GE must be participatory. Stakeholders in the intervention have a right to be consulted and to participate in decisions about what will be evaluated and how the evaluation will be done. In addition, the evaluation will assess whether the stakeholders have been able to participate in the design, implementation and monitoring of the intervention. It is important to measure the participation of stakeholder groups in the process as well as how they benefit from the results.

- **Fair Power Relations.** Both HR & GE seek, inter alia, to balance power relations between or within advantaged and disadvantaged groups. In addition, evaluators should be aware of the power of their own position, which can influence responses to queries through their interactions with stakeholders. There is a need to be sensitive to these dynamics.

Additionally, evaluators should preferably make use of mixed evaluation methods. Information from mixed methods can assist in the triangulation of data and increase reliability and validity, as well as being useful for exploring whether/why different stakeholder groups were affected differently by the intervention.

How are human rights and gender equality taken into account when preparing for an evaluation?

Before embarking on an evaluation, it may be worthwhile considering the evaluability of the HR & GE dimensions of the intervention. The process of assessing evaluability may entail the review of key documents and consultations with key stakeholders to capture the extent to which the HR & GE dimensions were incorporated in the design and implementation of the intervention.

Evaluators will often find different levels of evaluability in terms of HR & GE dimensions in the interventions to be evaluated. The UNEG handbook provides guidance on how to integrate these elements, regardless of whether evaluability is found to be low, medium or high.

- When evaluability is high, the HR & GE issues have been considered and are visible in all features of the intervention design, including in the contextualized intervention theory and intervention logic (log frame, indicators, activities, M&E systems). The design has benefited from HR &GE analyses and active stakeholder participation. When evaluability is high the role of evaluators/evaluation managers may be to ensure that the HR & GE dimensions are fully reflected in the evaluation terms of reference. It may also entail determining, with stakeholders, whether all areas are adequately covered and whether new methods and tools need to be introduced to capture any changes in intervention context and circumstances.

- Medium evaluability signifies certain coverage of HR & GE issues. It may be mentioned in various aspects of the design and intervention logic, but may not be fully articulated, or inclusion may be limited to a few disaggregated indicators (such as number of men and women). Important stakeholder groups may not have been included and there may be limited data on HR & GE issues in implementation and activity records. In addressing medium evaluability the evaluators/managers may seek to understand the reasons for such limitations, and the consequences for the programme implementation and results, highlight them in the evaluation terms of reference and include tools and methods in the evaluation design that would generate new information on HR & GE issues, and strengthen stakeholder participation.

- Low evaluability is indicative of scenarios where the intervention theory failed to consider the HR & GE dimensions in its design, implementation, monitoring and reporting. It entails an absence of disaggregated data and the participation of relevant stakeholder groups. When evaluability is low, evaluators/managers may reconsider all essential design features for the inclusion of the HR & GE dimensions, including why they were not covered, how this can be addressed in the evaluation terms of reference, how a relevant stakeholder analysis can be included and what the key data sources would be. The evaluation may focus on how lack of HR & GE perspective can compromise results. Provision should also be made for assertive recommendations, addressing the HR & GE dimensions, to be included in the final evaluation report.

An evaluability assessment may be a distinct exercise that is conducted well ahead of the evaluation, in which case it may also contribute to improving the design of an intervention in terms of its integration of HR & GE dimensions. Alternatively, it may be conducted in closer proximity to the evaluation and thus be a lighter process. The actual type of assessment would depend on the context. The UNEG handbook will serve to help evaluators/managers to make this judgement.

How are human rights and gender equality integrated into the evaluation terms of reference?

Once the evaluability assessment has been completed, the evaluation terms of reference can be prepared. A number of processes are typically involved, and HR & GE dimensions should be applied in each of them:

- including the HR & GE dimension as an explicit purpose/objective of the evaluation;
- identifying and engaging the stakeholders who will participate, and determining how they will participate;
- including the HR & GE dimension in standard evaluation criteria;
- framing the evaluation questions;
- selecting the indicators that will be used; and
- selecting the evaluation team.

Stakeholders

Stakeholders who are duty-bearers or rights-holders have special interests or responsibilities in the intervention. Evaluations should strive for the participation of both groups to ensure inclusion of balanced and diverse perspectives.

Duty-bearers may include government entities; officials; leaders; funding agencies; and those responsible for planning, funding or implementing the intervention being evaluated. Rights-holders may include groups and individuals ultimately affected or excluded by the intervention, disaggregated by age, sex and other relevant parameters; and other organizations with interests in the intervention or its outcomes, including women's organizations and other civil society organizations.

It is important to consider the role that each group of stakeholders might play in the evaluation, the gains from their involvement, the stage of the evaluation at which they can most usefully be engaged, and the ways in which they might be able to participate and their needs assessed in order to inform the evaluation. The principle of inclusion should guide the analysis. The UNEG handbook provides a useful HR & GE Stakeholder Analysis Matrix to guide evaluators/managers in making such decisions. Examples of questions in relation to stakeholder participation are:

- Beneficiaries, implementers, rights-holders and duty-bearers can be involved in the process with varying degrees of intensity. What will be the implications in terms of effort, time-line and budget?

- Is there a clear communication strategy with all stakeholders regarding who will participate, who will be consulted and who will make decisions when there are differences of opinion?

- Have the gains in credibility of the evaluation results from a particular level of participation been considered?

Evaluation criteria

The evaluation criteria commonly used in the UN are those developed by the Development Assistance Committee of the Organization for Economic Cooperation and Development (OECD-DAC) – relevance, effectiveness, efficiency, impact and sustainability. The UNEG handbook provides examples of the ways in which HR & GE dimensions can be integrated into each of these criteria. For example, under the relevance criterion the evaluator might consider the extent to which

the intervention complies with and contributes to the relevant international and regional conventions, declarations and international agreements, and the extent to which it was informed by analyses of human rights and gender issues and/or the needs and interests of diverse groups of stakeholders.

Evaluators might also consider whether criteria derived directly from HR & GE principles can be applied in the evaluation. These principles include equality, participation, social transformation, inclusiveness and empowerment.

Evaluation questions

The questions to be answered from the evaluation must be aligned with the evaluation criteria and must relate to each stage of the intervention, from design and planning, through implementation to results achieved. Different interventions and different evaluations will obviously require different questions, but a matrix in the UNEG handbook suggests some basic questions that may be a good starting point for the evaluator seeking to ensure that HR & GE dimensions are integrated meaningfully into his or her work. For example, under the effectiveness criterion, the evaluator might ask:

- Design and planning: Did the intervention's theory of change incorporate HR & GE dimensions?

- Implementation: During implementation, were there systematic and appropriate efforts to include various groups of stakeholders, including those who are most likely to have their rights violated?

- Results: What were the main results achieved by the intervention towards the realization of HR & GE?

Specifying these questions in some detail will also enable evaluators to identify the type of information that will be needed in order to answer them.

Indicators

Formulating indicators of HR & GE will assist the evaluator to assess the progress made on those dimensions as a result of the intervention and to identify the beneficiaries. The handbook offers advice for formulating both qualitative and quantitative indicators in the context of a particular intervention, and includes a number of examples drawn from specific types of interventions. A quantitative indicator of empowerment, for example, is the proportion of women and men in different stakeholder groups in decision-making

positions in local, national and sub-national government. A qualitative indicator is the change in access to information about claims and decisions related to human rights violations. More detailed examples are included in an Annex to the UNEG handbook.

Evaluation team

If the evaluation is to address HR & GE in an effective way, then the evaluation team must have the knowledge and commitment to apply these perspectives. While specialist expertise will be invaluable if it can be accessed, every member of the team should understand the UN mandates on HR & GE, and their application, and be committed to their inclusion in the evaluation. The *UNEG Ethical Guidelines* and *Code of Conduct for Evaluators in the UN System*[6] provide guidelines on ethics and behaviours for evaluators that are aligned with these approaches and are a further means of ensuring inclusiveness. They are reproduced in the UNEG handbook for the benefit of evaluators, along with desirable attributes for competence to integrate HR & GE within an evaluation team. Consider how diversity in the evaluation team can ensure a multiplicity of viewpoints and inclusivity.

How are human rights and gender equality integrated during the evaluation itself?

Integrating HR & GE dimensions is just as important throughout the conduct of the evaluation – the heart of the evaluation process – as it is in the planning phase. The UNEG handbook therefore suggests careful consideration when:

• selecting the evaluation methodology;

• collecting and analysing data;

• preparing the evaluation report; and

• disseminating the evaluation findings and developing a management response.

Selecting appropriate methodology

While all evaluation methodological designs are conducive to the integration of HR & GE dimensions, the use of a mixed methods approach is more likely to generate robust and accurate data on

---

6   See http://www.unevaluation.org/ethicalguidelines and http://www.unevaluation.org/unegcodeofconduct

the extent to which HR & GE were integrated in an intervention by allowing different perspectives to be heard, including those usually marginalized. For example, an HR & GE stakeholder analysis conducted in the preparation stage can help evaluators to select appropriate tools to maximize the participation of traditionally vulnerable and/or marginalized groups who may not normally be consulted during evaluation processes.

Other aspects to be considered in developing an HR & GE responsive methodology include ensuring an adequate sample that is inclusive of both women and men from diverse stakeholder groups; ensuring that data collection instruments allow for collection of disaggregated data; and, ensuring triangulation of data by collecting information from both rights-holders and duty-bearers.

Collecting and analysing data

HR & GE dimensions can be integrated in commonly used data collection and analysis methods by including HR & GE questions in the data collection tools and conducting an HR & GE analysis of the data collected. However, it also involves considering these dimensions in the process of collecting the data.

For example, evaluators organizing a focus group discussion (FGD) should consider in advance any barriers to participation when making decisions about the timing and location of the FGD and the composition of the group itself. Questions that evaluations might ask themselves include: What are the power dynamics within the group? Will women feel comfortable to speak freely if men are also present in the room? Evaluators can also ensure that other common tools, such as surveys, are in the format and language most appropriate for each stakeholder group. In some cases, surveys may need to be modified specifically for each group.

Preparing the evaluation report

Reports give evaluators the opportunity to highlight the importance of integrating HR & GE dimensions in an intervention, and the shortcomings derived from not doing so, and to illuminate the challenges and lessons learned and provide a clear explanation of the limitations or obstacles faced in integrating HR & GE in the design and implementation of the intervention, and in the evaluation process itself. This can help to stimulate future improvements and encourage actions to address limitations. Including specific recommendations on HR & GE dimensions will highlight specific areas for action by management.

Integrating HR & GE in evaluation reporting involves not only the inclusion of substantive elements in the report, but also thoughtful decisions on the most suitable forms of reporting. Reporting can move beyond the traditional written form and involve consideration of the different audiences for the findings and their specific information needs. Evaluators should ask themselves:

- How can evaluation findings be made accessible and understandable to both duty-bearers and rights-holders?

- Will different products need to be developed to ensure access to information?

- Has attention been paid to language and images to avoid stereotyping?

It may be necessary to make use of alternative forms of reporting, such as video, or to translate the report into relevant languages to ensure access to the relevant groups in order to avoid further marginalizing or disadvantaging them. Key findings may also be highlighted for evidence-based advocacy.

Dissemination and management response

Agency policies normally guide evaluation managers in the UN in developing dissemination strategies for evaluations. The integration of HR & GE in the process requires efforts to provide barrier-free access to the findings through translation, printing of hard copies and making use of relevant dissemination channels, among others, to reach both duty-bearers and rights-holders and direct and indirect users of the evaluation. These include human rights or gender organizations that may not have been involved in the intervention but for whom the findings would be of interest and use.

A management response which addresses the HR & GE recommendations, and provides action points in response, is crucial for enhancing the use of evaluation findings related to HR & GE issues, and to ensuring that the learning on HR & GE is incorporated into future practice. While normally developed by management, diverse stakeholder participation (including duty-bearers and rights-holders) in the process of developing the management response is one way of incorporating HR & GE dimensions into this last phase of the evaluation.

Conclusion

The development of guidance on the integration of HR & GE in evaluation fills a gap identified by UNEG. It enables evaluators to meet the expectations of the UN system to incorporate these dimensions into their work. This in turn will contribute to ensuring that HR & GE dimensions inform all aspects of the UN's activities from policy analysis to programme planning, implementation and monitoring and evaluation. The UNEG handbook constitutes an authoritative reference for UN evaluative work and is of relevance for the wider evaluation community. The handbook will be further refined on the basis of user feedback.

References

OHCHR, *UN common understanding of the Human Rights Based Approach to Programming.*

UN (2008), *Report on indicators for promoting and monitoring the implementation of human rights.*

UN ECOSOC (1997), *Resolution on Gender Mainstreaming.*

UNEG (2011), *Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance.*

WHEN HUMAN RIGHTS IS THE STARTING POINT FOR EVALUATION

Donna M. Mertens, Department of Educational Foundations and Research, Gallaudet University

"Citizens of former colonial powers are often baffled as to why indigenous or colonized peoples seem to suffer disproportionately from alcoholism, homelessness, mental illness, disease, lethargy, fatalism, or dependency. They cannot fathom… why many of their children cannot stay in school, or why many do not thrive in the contemporary, industrialized world of big cities and corporate capitalism. They are surprised that their development programmes don't produce the desired results and their attempts to alleviate the conditions under which so many indigenous or colonized peoples suffer may meet with passivity, indifference, resistance, or sometimes hostility."

Nobel Peace Prize winner
Wangari Maathai (2010, p. 172)

The United Nation's Universal Declaration of Human Rights (1948) stands as testament that the pursuit of human rights and the furtherance of social justice are the life-blood of this international organization. This declaration established that all people have a right to life; liberty; security of the person; equal protection under the law; freedom of movement; marriage with the free and full consent of the intending spouses; ownership of property; freedom of thought and religion; freedom of opinion and expression; peaceful assembly; participation in governance; work in just and favorable working conditions and education. Two points of tension are immediately apparent when considering this universal declaration of rights. Firstly, all people do not live in conditions in which these rights are afforded to them, as witnessed by world-wide events of resistance to repression and discrimination. Secondly, throughout the years since 1948, the UN has recognized that a universal declaration did not result in universal access to rights. Their recognition of the continued need for attention to human rights reflects awareness that specific groups of people suffer the denial of their rights more than others.

To remedy these disparities between people of privilege and those who are marginalized, the UN passed additional conventions and

declarations aimed at recognizing the most marginalized groups, including racial groups who are discriminated against (UN, 1969); people with disabilities (UN, 2006a); women (UN, 1979); children (UN, 1990a); migrant workers and their families (UN, 1990b); and, indigenous peoples (UN, 2006b). These declarations and conventions provide a partial list of the bases for discrimination and oppression that require focused attention if we are indeed to have universal human rights. The adoption of the Millennium Development Goals, and a review of progress toward meeting those goals, reinforces the need to attend to the poorest and most marginalized populations, including those who live in remote areas or urban slums, those who represent ethnic or racial minorities, and members of religious groups that experience discrimination (UNICEF, 2010).

The broader international development community has responded by calling for programmes that explicitly address the needs of the poorest and most marginalized communities, with conscious attention to the full spectrum of their diversity (Mertens, 2009; 2010; Mertens & Wilson, in press). This is evident in UNICEF's call for equity focused approaches in programme decisions (UN, 2010b), and a human rights approach to evaluation (Segone, 2009). On UNwomen (formerly UNIFEM) also supports the need to integrate human rights approaches with evaluation strategies, which focus on gender equity (Sanz, 2009). The UNDP Evaluation Policy was revised in 2011 to reflect a human rights focus:

"Evaluation is guided by the people-centred approach of UNDP to development, which enhances capabilities, choices and rights for all men and women. Evaluation abides by universally shared values of equity, justice, gender equality and respect for diversity" (p. 3).

This chapter addresses the challenges of planning, implementing and using evaluations that emerge when human rights is the starting point for policymakers, funders, programme developers and evaluators. It also takes on the difficult territory associated with gathering wisdom across small-scale evaluations to provide a human rights grounding for national level policymaking. Situating oneself as an evaluator in a human rights position requires re-thinking how evaluation is conceptualized, practiced, and used in international development. I make the argument that this re-framing of evaluation from this starting point based on human rights provides fertile ground for obtaining meaningful answers to questions about the efficacy of international development interventions.

Transformative paradigm

The transformative paradigm provides one philosophical framework that provides guidance for practical methods in evaluation, which align the human rights mission of international development organizations with a human rights approach to evaluation. The transformative paradigm provides a metaphysical umbrella to guide evaluators, which is applicable to people who experience discrimination and oppression on whatever basis, including (but not limited to): race/ethnicity; disability; immigrant status; political conflicts; sexual orientation; poverty; gender; age; or the multitude of other characteristics that are associated with less access to social justice. In addition the transformative paradigm is applicable to the study of the power structures that perpetuate social inequities. Finally, indigenous peoples and scholars from marginalized communities have much to teach us about respect for culture and the generation of knowledge for social change. Hence, there is not a single context of social inquiry in which the transformative paradigm would not have the potential to raise issues of social justice and human rights (Mertens, 2009, p. 4). Thus, the philosophical assumptions of the transformative paradigm serve to address the interests of not only the most deprived groups, but also to interrogate the power structures that can either support the attainment of human rights or can serve to sustain an oppressive status quo.

The transformative paradigm builds on the early work of Guba and Lincoln (2005), in defining the philosophical belief systems that constitute a paradigm in the evaluation context. They proposed that a paradigm was made up of four categories of philosophical beliefs:

1. *The axiological assumption* concerns the nature of ethics.

2. *The ontological assumption* concerns the nature of reality.

3. *The epistemological assumption* concerns the nature of knowledge and the relationship between the knower and that which would be known.

4. *The methodological assumption* concerns the nature of systematic inquiry.

The philosophical assumptions associated with the transformative paradigm provide a framework for exploring the use of a human rights lens in evaluation (see Figure 1).

Figure 1. Transformative philosophical beliefs

Axiological Assumption:

- Identification and respect for cultural norms that support human rights and social justice;
- Identification and challenge of cultural norms that sustain an oppressive system;
- Reciprocity – what evaluators give back to the community;
- Resilience – recognition and validation of the knowledge, expertise, and strengths in the community;
- Sustainability – facilitating conditions such that actions to continue to enhance social justice and human rights are feasible once the evaluator leaves the community;
- Recognition of limitations: Not over-stepping the evaluator's boundaries or over- promising.

Ontological Assumption:

- Recognizes that different versions of reality exist;
- All versions of reality are not equal;
- Recognizes privilege given to what is perceived to be real, based on: social, political, cultural, economic, ethnic, gender, religion, and disability perspectives;
- Interrogates versions of reality that sustain oppressive systems;
- Makes visible versions of reality that have the potential to further human rights.

Epistemological Assumption:

- Establishment of an interactive link between the evaluator and stakeholders;
- Acknowledges that knowledge is socially and historically located;
- Explicit acknowledgement of power inequities; and
- Development of a trusting relationship.

Methodological Assumption:

- Evaluators need qualitative assessment and dialogue time in the beginning of their planning in order to ascertain the cultural context in which they are working;
- Qualitative and quantitative data facilitate responsiveness to different stakeholders and issues;
- Methods used need to capture the contextual complexity and be appropriate to the cultural groups in the evaluation;
- A cyclical design can be used to make use of interim findings throughout the evaluation study; and
- Follow-up is needed to facilitate use to enhance the potential to achieve the strengthening of human rights.

Based on Mertens (2011)

Transformative axiological assumption

The transformative axiological assumption is discussed first because it provides the foundation for the subsequent assumptions. As evaluators reflect upon their beliefs about the nature of ethics, they first need to identify those ethical principles that guide their work. If they determine that the furtherance of human rights and the pursuit of social justice are the underlying ethical principles, then their beliefs are reflective of the transformative axiological assumption. Adopting such a position brings with it the need to consider differences of power and culture also as ethical issues. The questions then become: How can this evaluation contribute to social justice and human rights? The corollary being: What do I do differently in the evaluation in order to act upon these ethical principles?

A first methodological implication that emanates from this ethical stance is the need to identify the cultural norms and beliefs that are present in the targeted communities. In order to engage in culturally respectful ways, we need to include mechanisms for entering communities that permit identification of these norms and beliefs, and to understand the implications of those norms, either to support the pursuit of human rights or to sustain an oppressive system. For example, the African tradition related to "cleansing" a woman when her husband or child dies by bringing in a man from another village to have sexual relations with her is a practice that sustains an oppressive system and results in the continued spread of HIV/AIDS. This is a tradition that needs to be understood and challenged as sustaining an oppressive and dangerous practice. On the other hand, the African belief in Ubuntu ("I am because we are") serves as a belief that can support this challenge and lead to the Africans resolving the conflict between their beliefs in favor of that which is more humane.

What is the role of the evaluator in such a context? The transformative evaluator will identify, through respectful interactions, those cultural norms, that are supportive of human rights and those that support an oppressive system. The evaluator can engage with communities by arranging for culturally appropriate opportunities to address those norms and beliefs that conflict with the pursuit of human rights. They can insure that all stakeholder groups, especially those who are poorest or most marginalized, have a platform for authentic engagement with the evaluator. They can clarify that part of the conducting of an ethical evaluation is inclusion of the goal

to leave the community better off than before the evaluation was undertaken. This means that the evaluator proactively addresses issues of reciprocity, sustainability and how the evaluation can contribute to the possibility of taking action to enhance social justice and human rights after the evaluator leaves the community.

As Maathai (2010, p. 130) noted:

"Poor people need to be engaged in their own development, and, by extension, in expanding the democratic space that many African societies desperately need. Just as communities should be mobilized to combat malaria, or HIV/AIDS, for instance, so they must work together to fight the scourges of failed leadership, corruption, and moral blindness. However, because the poor are more likely to be uneducated, illiterate, and ignored, and to feel powerless, this requires both political and economic commitment, as well as patience and persistence, since change does not occur overnight".

Although Maathi's comments are situated in the African context, they have relevance for the broader communities of poor and marginalized peoples. In such communities, the people may have lost sight of their strengths as a community. The evaluator can work with the people to explore their strengths and to validate their knowledge as having value because it is rooted in an experience that few evaluators and policymakers have had. This axiological assumption has specific implications for the transformative ontological assumptions.

Transformative ontological assumption

As alluded to in the previous paragraph, people with different experience-bases will often have different perceptions of reality. Thus, the transformative ontological assumption leads the evaluator to recognize that different versions of reality exist and that all versions of reality are not equal. The transformative evaluator has the responsibility of uncovering the different versions of reality that exist and to interrogate the basis for privileging one version of reality over another on the grounds of different social; political; cultural; socio-economic; gender; age; religion; geographic; and, disability perspectives. Just as different beliefs and norms are associated with either support for, or inhibiting human rights, the versions of reality need to be examined on the same basis. For example, if people in power perceive that it is too expensive to try to address the needs of the poorest and most marginalized people, then it is not

likely that the poor people will receive the support they need for a good quality of life. However, a version of reality that holds that on moral grounds alone (and perhaps on economic grounds as well, see UNICEF, 2010) we have an obligation to address the needs of the poor and marginalized, holds the potential to actually further the human rights agenda.

From the transformative ontological stance, the evaluator needs to ask questions about the quality of the evaluation that focus on the extent to which the evaluation reveals the different versions of reality, and the consequences of accepting one version of reality over another in terms of furthering human rights. The evaluator needs to document the fact that different versions of reality were explored, and that the consequences of those versions were considered in terms of the evaluation's contribution to the needed social changes.

Transformative epistemological assumption

In evaluation, the epistemological assumption can be thought of in terms of the nature of the relationship between the evaluator and the stakeholders (rather than the philosophical language about the relationship between the knower and that which would be known). In order to act in line with the transformative axiological and ontological beliefs, the evaluator needs to build a relationship with the stakeholders, which is based on cultural respect, acknowledges power differences, is inclusive of marginalized voices, and provides a safe environment for everyone to express themselves. The evaluator should provide evidence that they have considered issues of power and that they have included the voices of the less powerful in an accurate way, which leads to appropriate social action. In order to do this, the evaluator needs to address issues of power and language and to build trusting relationships with the stakeholders. The nature of those relationships will be contextually dependent, as the cultural norms that dictate what will engender trust in one setting will differ in another setting.

Evaluators have resources at their disposal to ensure that they are aware of the relevant dimensions of diversity within a particular geographic area, as well as to become informed about the cultural beliefs and norms that have implications for the development of the relationships with stakeholder groups. In addition to the mainstream evaluation literature, members of marginalized communities have increasingly published documents that give insights into appropriate entry into and means to establish relationships with

members of their communities. Harris, Holmes and Mertens (2009) provide terms of reference for conducting evaluations in the Deaf Sign Language community. Chilisa (2009) and Ntseane (2009) provide guidance for the ethical conduct of evaluations in Africa. Cram (2009) describes the protocol for entering and engaging with Maori communities in New Zealand.

The common thread in these culturally specific examples is that evaluators, especially those who are not working in their native language or in their native culture, need to consciously address the challenge of entering the targeted community. If feasible, the evaluator can establish lengthy relationships with community members who demonstrate their willingness to understand the culture. However, with shorter term evaluations, evaluators need to:

"present themselves and their backgrounds in ways that make clear their strengths and limitations in terms of their knowledge and life experience. This positioning allows the evaluator to acknowledge the need to work together with the people from the community, who have a stronger understanding of cultural and social issues (Mertens, 2011, p. 6)."

Engagement with communities is key to the conduct of transformative evaluations. Additional strategies include hiring evaluators from the home communities, and establishing teams of evaluators who have expertise in evaluation and awareness of the culture of the community. This might entail capacity building activities for those who lack evaluation skills or for those who lack cultural skills. Evaluators can also form relationships with important community gatekeepers who can vouch for the evaluators' credibility. However, this strategy comes with a cautionary note: the evaluator must be cognizant of the dangers associated with accepting one or a few people as representatives of the voices of the larger group. They should also avoid token representation of stakeholders; the invitation for participation must be authentic. In keeping with the international community's movement toward country-led evaluations (Segone, 2009), the transformative approach to evaluation supports those relationships between evaluators and community members that acknowledge the power differences and that value the expertise brought-in by each team member.

Transformative methodological assumption

No single method is dictated by situating oneself in the transformative paradigm. Rather, the methodological assumption flows from the axiological, ontological, and epistemological assumptions in that the focus is on developing methods that facilitate the support for human rights, uncover competing versions of reality and the consequences of accepting one over the other, and establishing trusting relationships with stakeholders. The model of transformative methods that encompasses these principles is a cyclical design somewhat akin to Patton's Developmental Evaluation (Patton, 2011). However, Patton does not situate his work within a transformative philosophical belief system. Hence, differences between his approach and a transformative approach emerge because the latter starts with the principle of furthering human rights.

The transformative methodological assumption calls for a re-thinking of the conceptualization of evaluation currently operating in the international development community. The definition of evaluation within the international development community tends to focus on measuring the outcomes of projects. The international community distinguishes between monitoring (the continuous function that provides managers and stakeholders with regular feedback), and evaluation. The United Nations Development Programme (2011, p.4) defines evaluation as follows:

> "Evaluation is a judgment made of the relevance, appropriateness, effectiveness, efficiency, impact and sustainability of development efforts, based on agreed criteria and benchmarks among key partners and stakeholders. It involves a rigorous, systematic and objective process in the design, analysis and interpretation of information to answer specific questions. It provides assessments of what works and why, highlights intended and unintended results, and provides strategic lessons to guide decision-makers and inform stakeholders."

This definition, and others found throughout the international community, focuses primarily on evaluation as an act that occurs after a programme is implemented. However, the broader evaluation community sees evaluation as a strategy that has the potential to inform the development of programmes, provide feedback about the processes that are implemented, and document the intended and unintended outcomes and impacts of a programme. In order to address the human rights mission of international development organiza-

tions through a human rights evaluation approach, evaluators need to be involved in international development projects throughout the life of the project, from the moment of inception.

Thus, the transformative methodological assumption supports the use of a cyclical model that provides opportunities for community participation throughout the project's lifetime (Mertens, 2009; 2010). If evaluators are called to do their work only after a programme is implemented, and to focus their efforts on outcomes and impacts, then they have missed the many opportunities afforded by a cyclical evaluation approach: to contribute to the quality of a project; to allow for on-going adjustments needed to correct or enhance the project; to gather wisdom from the community in order to support this effort; and to document outcomes and impacts with validity.

Before there is a programme, or when changes in an existing programme are warranted, members of the community can be consulted about the need for a programme or for changes in an existing programme, and on what that new or revised programme might look like in order to be culturally responsive. This stage of the process can involve the review of documents; meetings with various stakeholder groups; review or development of culturally appropriate protocols for interaction; focus groups; 'town hall' meetings or indigenous equivalents; and/or surveys. In evaluation, this period of the evaluation would be termed a needs assessment or context evaluation. The results of this first round of data gathering need to be analyzed in conjunction with community members in order to provide culturally appropriate interpretations of the data, and to determine the next steps in the process. This use of information, to inform decision-making, has been present in the broader evaluation community for decades. If this approach is to be applied in international development, it would expand the conceptualization of evaluation as it currently stands, in that context.

Based on information from the first stage of the evaluation activities, the evaluators would then work with programme developers and community-based stakeholders to determine the nature of the intervention, as well as to specify the contextual variables that require attention in its implementation. This might include the development of a pilot version of the intervention, which is implemented on a smaller scale in order to examine the appropriateness of the intervention, the process of implementation, and possible measures of outcomes. Data from this phase of the evaluation can

be used to inform decisions about the intervention and its implementation, as well as to make adjustments to the evaluation measures, as needed. These data should be brought to the programme developers and community stakeholders on a continuous basis so that adjustments can be made as needed.

Based on the information from the second stage, the evaluators can then work with the programme developers, implementers, and community stakeholders to explore the process of scaling-up the intervention, still maintaining a focus on process evaluation (how is the programme being implemented?) and on the outcomes and impacts, with an awareness that adjustments may need to be made on the basis of heterogeneity in marginalized communities. For example, if an HIV/AIDS prevention programme is implemented in the hearing community, what needs to change to have a quality programme in the deaf community? This transfer of the programme to a different marginalized group, who may live in the same geographic area as the hearing community, requires involvement of a different group of community stakeholders.

Follow-up of the use of the evaluation findings for programme changes and for policy decisions is a crucial part of the transformative methodological assumption. An evaluator's work is not finished by the provision of a final report to the funder. Rather, they need to work with appropriate constituencies to facilitate the use of the information by engaging in conversations, focus groups, 'town hall' meetings or indigenous equivalents, and/or interviews with appropriate stakeholders to determine how they can use the information. This should be viewed as part of the evaluator's responsibility and comes under the concept of meta-evaluation. If the goal of the evaluation is to address human rights, then the follow-up to facilitate and document use needs to be included in the definition of evaluation. Such follow-up is in keeping with the spirit of the international community's desire to make sense of the multiple evaluations that are conducted in order to gather wisdom from them for the purpose of policymaking decisions at a broader level. Quality local evaluations are needed so that the tension between the attention to diversity and the need for broader evidence-supported policies is addressed.

Evaluators who situate themselves in the transformative paradigm use methods that allow them to capture the "contextual complexity and provide pluralistic avenues for engaging appropriately with diverse cultural groups in the evaluation" (Mertens, 2011, p. 7). Hence, many

transformative evaluators choose to use mixed-method approaches that allow for the combination of both quantitative and qualitative data in a cyclical manner, to inform decision-making throughout the lifetime of the intervention. Transformative mixed-method designs reflect the cyclical approach described earlier, and support the use of mixed-methods as a mechanism for engaging with the full range of stakeholders and providing the needed evidence of programme effectiveness.

At the beginning of the evaluation, evaluators begin with qualitative assessment and dialogue time, in order to ascertain the cultural context in which they are working. They also benefit from the collecting of quantitative data that allows for a broader sense of community strengths and challenges, as well as for documentation of the effectiveness of earlier interventions. During the pilot stage, case study approaches can be combined with counterfactual comparisons. During the implementation and possible scaling-up stage, both qualitative and quantitative measures can be used, along with designs that provide counterfactual comparisons, as long as the ethical considerations of denial of treatment are addressed adequately.

Conclusions

The international development community has been a strong supporter of human rights for decades, especially for the poorest and most marginalized populations. The international development evaluation community has taken the stand that their evaluations should align with this support for human rights. If evaluators take the promotion of human rights as their starting point, then they need a framework for thinking about the implications of this stance. The transformative paradigm offers such a framework through its examination of ethics and reality, and through the relationships between evaluators and stakeholders, together with evaluation methods that are rooted in the pursuit of social justice and the furtherance of human rights. The thinking of feminists; indigenous peoples; critical theorists and critical race theorists; disability rights advocates; and deafness rights advocates are commensurate with the transformative paradigm's philosophical assumptions. Practical implications for evaluators derived from these assumptions, provide guidance for evaluators who align their work with the international community's mission to address human rights issues. The international development evaluation community stands to benefit by accepting a transformative cyclical approach in order to reflect the complexities that

challenge programme effectiveness and, which provide a platform for informing policies that can enhance the possibility of achieving the desired end – a better quality of life for those who suffer discrimination and oppression.

References

Chilisa, B. (2009). Indigenous African-centered ethics: Contesting and complementing dominant models. In D.M. Mertens & P.E. Ginsberg, (Eds.), Handbook of social research ethics (pp. 407-425). Thousand Oaks, CA: Sage.

Cram, F. (2009).Maintaining indigenous voices. In D.M. Mertens & P.E. Ginsberg (Eds.), Handbook of social research ethics (pp. 308-322). Thousand Oaks, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluence. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 191–215). Thousand Oaks, CA: Sage.

Harris, R., Holmes, H., & Mertens, D. M. (2009). Research ethics in sign language communities. *Sign Language Studies, 9*(2).

Maathai, W. (2010). The challenge for Africa. London: Arrow Books.

Mertens, D.M. (2009). Transformative research and evaluation. NY: Guilford.

Mertens, D.M. (2010). Research and evaluation in education & psychology: Integrating diversity with quantitative, qualitative, & mixed methods (3rd ed.), Thousand Oaks, CA: Sage.

Mertens, D.M. (2011). Social transformation and evaluation. *Evaluation Journal of Australasia*. 10(2), 3-10.

Mertens, D.M. & Wilson, A.T. (in press). Program evaluation from theory to practice: A comprehensive approach. NY: Guilford.

Ntseane, P.G. (2009). The ethics of the researcher-subject relationship: Experiences from the field. In D.M. Mertens & P.E. Ginsberg (Eds.) *Handbook of Social Research Ethics*, pp. 295-307, Thousand Oaks, CA: Sage.

Patton, M.Q. (2011). Developmental evaluation. NY: Guilford.

Sanz, B. (2009). Gender equality and human rights responsive evaluations: Rethinking approaches. In M. Segone (Ed.), From policies to results, pp. 125-146. NY: UNICEF.

Segone, M. (Ed.) (2009). Country led monitoring and evaluation systems. NY: UNICEF.

UN Evaluation Group. (2005). Norms for evaluation in the UN system. NY.

United Nations Development Programme. (2011). The evaluation policy of UNDP. NY: Author.

UNICEF. (2010). Narrowing the gaps to meet the goals. NY.

UNICEF. (2010). Progress for children: Achieving the millennium development goals with equity. NY: Author.

United Nations. (1948). *Universal declaration of human rights.* Retrieved February 11, 2008, from *www.un.org/Overview/rights.html*.

United Nations. (1969). *The international convention on the elimination of all forms of racial discrimination.* New York: Author. Retrieved February 11, 2008, from *www.ohchr.org/english/law/pdf/cerd.pdf*.

United Nations. (1979). *The convention on the elimination of all forms of discrimination against women (CEDAW)*. Retrieved February 11, 2008, from *www.un.org/womenwatch/daw/cedaw/text/econvention.htm*.

United Nations. (1990a). *Convention on the rights of the child.* Retrieved February 11, 2008, from *www.ohchr.org/english/law/pdf/crc.pdf.*

United Nations. (1990b). *International convention on the protection of the rights of all migrant workers and members of their families*. Retrieved February 11, 2008, from *www.un.org/millennium/law/iv-13.htm*.

United Nations. (2006a). *Declaration of ad hoc committee on a comprehensive and integral international convention on the protection and promotion of the rights and dignity of persons with disabilities*. Retrieved May 10, 2008, from *www.un.org/disabilities*.

United Nations. (2006b). *Declaration of rights of indigenous peoples*. Retrieved March 21, 2007 from *http://www.un.org/esa/socdev/unpfii/en/declaration.html*

STRENGTHENING EQUITY-FOCUSED EVALUATIONS THROUGH INSIGHTS FROM FEMINIST THEORY AND APPROACHES[1]

Katherine Hay, Senior Evaluation Specialist, International Development Research Centre, Regional Office for South Asia and China

Introduction

Given the persistence of inequities globally, evaluation must play a stronger role in understanding how societies change and what policies and programmes show promise in shifting norms and inequities. This section examines how an equity lens and transformation-oriented practice can influence the framing, methods, and conduct of evaluation. The chapter draws on a set of evaluation efforts and experiences in India, and explores how a rights and equity lens can play out in evaluation practice. Most of the cases that are explored in this section focus on gender based inequities, a persistent site of exclusion that shapes the experience of all policies and programmes in India. However, ideas from these cases can be extended to other types of inequities. The section attempts to document and analyze experiences, perspectives, and ideas from practice and to connect them to emerging developments in Equity-focused evaluation.

The past decade has seen renewed enthusiasm and interest in evaluation in international development. Donors are putting out calls for demonstrating 'impact' and governments, including in India, have articulated interest in strengthening evaluation systems and capacity (Hay, 2010). Recently, India has announced plans to strengthen their evaluation system and to set up an independent evaluation office. With the expanding space in India for evidence, critique, and debate on development programming and policies, it is timely to explore the extent to which equity-focused approaches could inform evaluation questions, design, practice, and use.

---

1   Based upon and building on a paper in the Indian Journal of Gender Studies by the author (forthcoming).

This section argues that central ideas from feminist theory and research can strengthen and inform Equity-focused evaluation. The section starts from an understanding of inequities as manifest and systemic in social institutions. Equity-focused evaluation is presented as a way of understanding how intersecting social cleavages (such as gender, race, class, sexuality, caste, and religion) define and shape the experience and the exercise of power in different contexts. These ideas draw on work by Sudarshan, Ramachandran, Khanna, Jandhyala, and Murthy (forthcoming). Together with other social scientists and evaluators, between 2010 and 2011, they examined and reflected on their evaluation practice through a series of workshops and writing. Examining their evaluation experiences and reflections, this chapter explores how an equity-focused stance can influence the framing, methods, and practice of evaluation in India.

Conceptual framework

The definition of Equity-focused evaluations suggested by Segone in the introductory chapter of this book is consistent with recent definitions provided by Podems (2010) on feminist evaluation. Podems describes feminist evaluation as flexible and as being a 'way of thinking about evaluation.' Podems (2010) gives examples of practical aspects of feminist evaluations including their interest in multiple factors and structures influencing inequities, of which gender is only one. This can be compared to Segone's introduction: "inequity is rooted in a complex range of political, social, and economic factors that include, but are by no means limited to: gender discrimination" (Segone, 2011). Feminist theory offers deep and rich literature on the intersecting nature of exclusions (MacKinnon 2006, Mohanty 2003, Narayan 1997), which can inform evaluation practice.

Understandings of feminist evaluation, and of Equity-focused evaluation provided in this volume, are clear that documenting inequities is inadequate; quality evaluations using either approach should also seek to reduce those inequities. In doing so, both approaches implicitly recognize that evaluation can serve to reinforce or to challenge existing inequities. For example, Podems notes that while some evaluations might identify or record the differences between men and women, feminist evaluations would explore why these differences exist and challenge 'women's subordinate position' (p. 8). Though implicit, in both definitions it is recognition that for evaluations to change and challenge inequities they need to be used. This section makes that use explicit for Equity-focused evalua-

tion, as without use of evaluation, the purpose of equity cannot be achieved.

Drawing from the work of a range of evaluation theorists, Podems (2010) lays out 6 tenants of feminist evaluation (Box 1). It has as a central focus on gender inequities; it recognizes discrimination based on gender is systemic and structural; evaluation is political; knowledge has power; knowledge should be a resource of and for the people who create, hold, and share it; and there are multiple ways of knowing and some are privileged over others. If the focus on gender is expanded to include all sites of inequity, these tenants can arguably also be equally and usefully applied to all sites of inequity. In the case of UNICEF for example, the first tenant might include a central focus on inequities facing children marginalized by gender, disability, ethnicity or other sites of exclusion. Keeping the other tenants the same, and adding 'use' as a separate explicit tenant discussed above, what does this set of principles bring to equity-focused evaluation?

> ### Box 1: Six Tenants of Feminist Evaluation defined by Sielbeck-Bowent et al (2002) and compiled in Podems (2010):
>
> - Feminist evaluation has as a central focus the gender inequities that lead to social injustice.
> - Discrimination or inequality based on gender is systemic and structural.
> - Evaluation is a political activity; the contexts in which evaluation operates are politicized; and the personal experiences, perspectives, and characteristics evaluators bring to evaluations …lead to a particular political stance.
> - Knowledge is a powerful resource that serves an explicit or implicit purpose.
> - Knowledge should be a resource of and for the people who create, hold, and share it. Consequently, the evaluation or research process can lead to significant negative or positive effects on the people involved in the evaluation/research…
> - There are multiple ways of knowing; some ways are privileged over others.

Let us first explore the nature of these principles. The feminist lens brings an emphasis on power relations, structural elements of inequities, justice, and politics into evaluation. This lens and these foci can provide important and different perspectives on Equity-focused evaluations. Through these foci, feminist research has made important contributions to development more generally, including on such diverse topics as: women's work and the double work burden; social cleavages and overlapping sites of discrimination; the 'black box' of

the household; and understandings of rights that are less abstract and more lived. Specifically, just looking at the field of economics, the contributions of a feminist lens are extensive and important. Researchers have both used mainstream economic tools to examine wage gaps between men and women and critiqued these tools for their limited ability to shed light on the underlying inequities behind such gaps (Figart, Mutari and Power, 2002). Studies of unpaid work within households, the 'black box', have brought attention to women's unpaid work (Waring, 1988) and highlighted inequities in distribution of resources within households (Sen 1990). Tools developed by McElroy and Horney (1981) have become commonly used for understanding decision making and agency within households. A feminist lens has also lead to innovations in the analysis of government budgets according to their effects on gender equity (Budlender et al., 2002) and understanding the effects of macroeconomic policies of structural adjustment and liberalization (Çagatay, Elson and Grown, 1995; Grown, Elson and Çagatay, 2000). For example, such research has demonstrated that actions such as cutbacks in health care programmes often have their most immediate impact on women. Feminist economists have also analyzed how factors such as race and caste (Brewer, Conrad and King, 2002) interact with gender and affect economic outcomes.

Just looking at this very brief list of how feminist insights and analysis have strengthened one field of development (economics), one can draw parallels to the way in which such analysis can bring new developments, approaches and insights to Equity-focused evaluation. Equity-focused evaluation can focus attention on different variables, and in doing so, challenge or critique the dominant discourse underpinning the programme. Examples of the ways in which central ideas on the structural nature of inequities can be explicitly brought into evaluation design include:

- Collecting data on women's time and women's drudgery in income generating projects (Sudarshan and Sharma, forthcoming, Murthy, forthcoming);

- Examining which children from which social groups have access to better schooling and their experience within those school systems (Ramachandran, forthcoming);

- Documenting how elite women capture benefits from political decentralization processes to the detriment of other less privileged women or men (Devika and Thampi. 2010).

Feminist work on breaking into the 'black box' of the household, understanding and examining overlapping sites of inequities, and bringing focus to women's work and women's double burden, creates rich terrain that Equity-focused evaluation could and should more explicitly draw from.

With the modifications suggested, it is suggested that tenants from feminist evaluation (themselves drawn from a broader field of feminist research in multiple disciplines) can be usefully applied to all Equity-focused evaluations to bring useful and needed insights into Equity-focused evaluation. The rest of this section explores and suggests what this may entail in practice.

Evaluation is a process. Different evaluation theorists and practitioners categorize it into different stages, but in general: there is a start or a planning phase that includes deciding what to evaluate and what questions to ask; a design phase of determining what methodologies and methods will best generate the kind of knowledge and evidence needed; an implementation phase where data is gathered and analysed, and; a phase of use where the evaluation findings are shared and applied. These phases usually overlap in different ways depending on the nature of the evaluation.

This chapter applies a modified set of principles drawing from feminist analysis and utilization-focused approaches to evaluation stages (see figure 1) to explore through a number of case examples, how researchers and evaluators can integrate these principles into their Equity-focused evaluation work. The following sections look at how the ideas can influence the framing of evaluation questions, evaluation design, evaluation judgements, evaluation practice, and evaluation use.

Figure 1: Principles for Equity-focused evaluation drawn
from feminist research and approaches and stages in
the evaluation process.

Principles:

1. Has a central focus on inequities

2. Recognizes that inequities are structural

3. Recognizes that evaluation is political

4. Recognizes and values different ways of knowing

5. Proposes to add value to those who are marginalized

6. Requires use

Evaluation Questions

Evaluation Design

Evaluation Judgements

Evaluation Practice
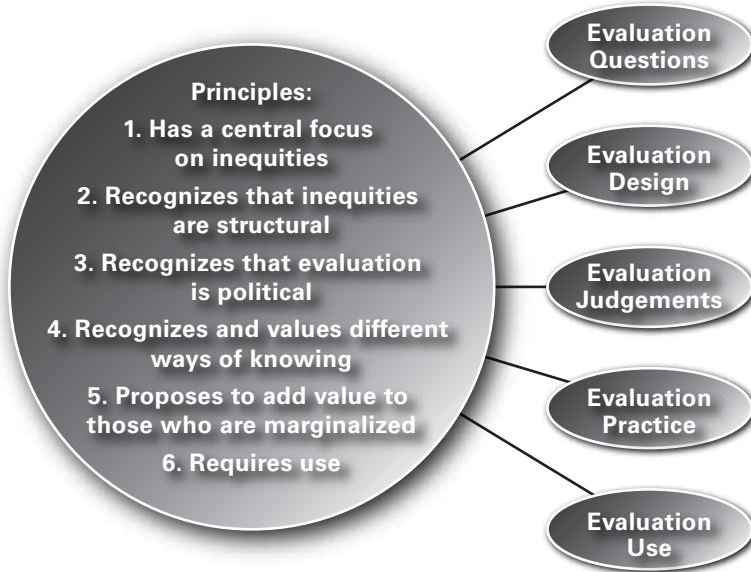
Evaluation Use

Setting evaluation questions

Programme theory is informed by competing discourses on development and equity that are at times explicit, more often implicit, and at times competing. Integrating the analysis of power and the structural nature of inequities into Equity-focused evaluation offers opportunities to critique dominant discourses and hold them up for scrutiny. 'Discourse' here means the 'big ideas' that shape our understanding of how the world works. Discourse matters because it underpins and legitimizes interventions. For example, policy responses to HIV were based on dominant discourses on HIV that evolved over time. These discourses included 'gay plague' discourse, a 'contaminated other' discourse, an 'innocent victim' discourse, a 'heterosexual-risk' discourse and…a 'development' discourse on AIDS (Hill, 1995). Discourses can (sometimes) be unpacked fairly easily in hindsight but can often be obscure (often intentionally so) as they are lived. Ideas from deconstructive analysis (Dietz 2003, Nash 2002) from feminist research can provide useful starting points to ground this effort to unpack discourse, as can realist synthesis approaches from within the evaluation field (Pawson, 2006). Evaluation can raise questions about the discourse

itself, the way it is articulated in policies and programmes, and whether the implicit theories around the nature of the problem, and how change will happen, hold true on the ground.

Trends in discourse evaluators using equity-focused approaches should consider the dominance of questions on impact and increasingly, with some funders, on value for money. At times these are the most critical or important factors to examine. At times they become political rallying calls, which are inserted into evaluations rather thoughtlessly. In such cases, evaluators should demand a more thoughtful discussion on the use and users of the evaluation and negotiation to change the questions when it is not possible to measure impact or cost, or where it is not the most important issue. For example, evaluators describe shifting the questions of impact and attribution to new questions on assessing 'how effective are strategies in particular contexts' or whether outcomes are in line with needs of beneficiaries (Sudarshan and Sharma, forthcoming). Equity-focused evaluation may also include valuing, and thus generating knowledge on process results and unintended outcomes. Reflecting on several evaluations of gender programmes and organizations, Sudarshan and Sharma (forthcoming) note that, in their experience, a more 'iterative framework' and approach to evaluations is most useful for capturing unintended outcomes. Several evaluation theorists (Morell 2005, Mertens 2009) have made this a focus for their work. However, it is particularly important in evaluations relating to structural inequities, as interventions may further reinforce inequities in ways that were not anticipated, or in attempting to shift those inequities may create conflict or reinforce other divisions.

Ramachandran (forthcoming) illustrates how changing discourse around education in India in the last 50 years has influenced programmes; from education being conceptualized as a 'universal good' at the time of independence (1947), to an instrument for population control in the 1960s, to a 'right' by the 1980s, and to a cornerstone of women's 'empowerment' by the 1990s. Evaluation can be used to examine the ways in which dominant discourses become lodged in policies and programmes and test whether the implicit assumptions behind these discourses resonate with the actual experience of marginalized groups on the ground. In doing so, evaluation offers opportunities to reshape and critique the discourse informing those policies, and to bring a greater diversity of views and values into that discourse. Equity-focused evaluations can ask, 'who has constructed this discourse and whose experiences are not reflected?'

Jandhyala's (forthcoming) account of the Mahila Samakhya pro-gramme, a large programme on women's education and empower-ment in India, reflects the understanding of evaluation as a political space where competing discourses that inform programme theory can be examined. She described how an external donor coming into the programme made funding contingent on defining results and tar-gets in a way that reflected an understanding of the changes sought that was fundamentally different from that of the implementing organization. The funder wanted to show progress towards achiev-ing the Millennium Development Goals, with women's mobilization through the programme being a means to reach goals around ele-mentary education. The target associated with Millennium Develop-ment Goal three is to eliminate gender disparity in education. So the implicit change theory is that more educated and informed women will be more likely to send their daughters to school. This vision however, is quite different from that of the programme implement-ers, who saw women's mobilization and empowerment as the goal itself. The monitoring and evaluation framework became the space where different views on the nature of structural inequities and the 'goals' of empowerment were articulated.

Evaluation can also identify gaps in programme theory that weaken opportunities to address inequities. For example, India has a set of programmes that has some parallels with depression era works programmes of the United States, including a huge programme designed to create jobs through building infrastructure in rural areas, called the Mahatma Gandhi National Rural Employment Guar-antee Scheme (MGNREGS). Analysis of this programme, illustrates that the design of the programme was gender sensitive in a range of ways (equal wages for men and women, participation of women in committees, provision of a crèche on work sites etc.). Evalua-tions of the MGNREGS, illustrate how evaluations can test different dimensions of programme theory (Sudarshan and Sharma, forth-coming). However, evaluations showed not only that crèches are often not set up – an implementation failure – but also that women workers were more comfortable leaving infants in the care of older children – a programme theory failure. The second finding speaks to the design itself and the ideas informing that design. Specifically, it is based on an assumption that parents would choose to leave their children at crèches rather than with family members. By get-ting both the design and the implementation wrong the programme led to older female children missing school in order to provide child-care.

There is an opportunity for evaluators to identify and test the theories underpinning policies and programmes – including by looking across sets of programmes. For example, in the case above, how do assumptions about gender roles and child care feed into development programmes and schemes that cut across a range of development domains? How do assumptions of women and girl's labour preferences and time inform (if at all) discourse on development plans? To what extent do programmes assume women's time is unlimited and unintentionally move domestic burdens onto girl children by further engaging women in productive and development work?

Implicit (or even explicit) theoretical underpinnings of programmes can vary among actors and can change and shift over time. The space to explore and critique this may be very limited. As Khanna (forthcoming) notes, "Unequal power relations are so deeply internalised within hierarchical bureaucracies that discussion on gender power relations is next to impossible." Nonetheless, evaluation designs and frameworks can serve to embed, ignore, reflect or challenge those underpinnings and relationships. A recognition of evaluation as a political space can bring these tensions to the surface and promote more transparent review and dialogue on competing or alternative values or theories.

Evaluation design

Equity-focused evaluation can be situated within different approaches to evaluation and draw upon the differing traditions in design, methodology, and approaches to rigour and validity found in those approaches. As Segone notes, "most of the Equity-focused evaluation data collection and analysis techniques is built on approaches with which many practitioners in development evaluation already have some familiarity: the emphasis is on refining and refocusing existing technics – and enhancing national capacities to use those technics – rather than starting with a completely new approach" (Segone, 2011).

Equity-focused evaluation is not one design or one set of methods, but a lens or standpoint that influences the choices made in design and methods. A rigorous Equity-focused evaluation would be one that used the range of methods that best matched the questions around the type of change the policy or programme is addressing. In this view – equity is the lens in which questions are asked and evidence is challenged. Feminist analysis brings to Equity-focused evaluation, recognition that this process is constructed and political.

Individual methods *per se* are not 'equity-focused or feminist'; their suitability (and rigour) in any given evaluation is a function of their ability to generate valid and reliable data that speaks to the nature and change around the inequity that the programme is attempting to address. That said, there are a range of lines of evaluation and research theories that focus on issues of equity, rights and voice; for example, the transformative evaluation paradigm (Mertens 2010) and participatory paradigms (Chambers 1983, 1987). Equity-focused evaluation (and empowerment, feminist, participatory and transformative approaches) must start with the principle of includ-ing the voice of the stakeholder, as do these approaches to evalu-ation. However, what is considered 'valid and reliable' in Equity-focused evaluations often mirrors broader trends in evaluation and development.

On-going debates on qualitative versus quantitative approaches remain, but are now superimposed with other new debates on experimental (randomized) and quasi-experimental designs. Despite debates that are at times polarizing, there does appear to be a broader openness and interest in more mixed-method approaches and rec-ognition that different designs suit different questions, contexts, and resources. Ramachandran (forthcoming) points to the impor-tance of mixed methods using an example from education evalua-tion. Detailed observation-based studies revealed that parents were sending more boys to private schools and more girls to government schools. These qualitative micro studies, in turn, created demand for larger quantitative data sets, on both government and private school admissions (disaggregated by social group and gender) (Ramachan-dran, forthcoming). With the passage of the Right to Education bill in 2010, India moved towards capturing this data through larger quanti-tative data sets but it was the smaller more qualitative studies that raised the issue, and which continue to be the only source of evi-dence on this issue of exclusion and discrimination.

Evaluators conducting Equity-focused evaluations must recognize the perceptions of credibility by intended users of some evaluation designs (and of quantification more generally). As noted earlier, without use of evaluation, the purpose of equity cannot be achieved and use can at times be centred on perceived methodological cred-ibility. This can also make choice of method a strategic choice. For example, Sudarshan and Sharma (forthcoming) note:

"... the current dominant mode in evaluation design emphasizes quantification stemming from the need to provide evidence of

success to the donor, and the fact that numbers are far more effective in advocacy than narratives are…"

Different types of knowledge, expressed through different methodological traditions, have more, or less, power in decision-making structures. Noting that women's grass roots and implementing organizations in India, 'shy away from quantitative and macro data,' Khanna (forthcoming) integrates quantitative evaluation in her evaluations "… to increase their mastery over quantitative data." The feminist lens brings recognition of the power of quantitative data and the transformative potential of empowering organizations working with marginalized groups with the capacity to use both quantitative and qualitative data through the Equity-focused evaluation process itself.

Evaluation judgments

The purpose of evaluation includes improving the accountability and compliance of programmes and organizations, and knowledge development (Mark, Henry, and Julnes, 2000). Making a judgement about what works and what does not work is also a fundamental purpose of evaluation. As Segone notes in the introduction, "An Equity-focused evaluation is a judgment made of the relevance, effectiveness, efficiency, impact and sustainability of policies, programmes and projects on equitable development results" (Segone, 2011).

Much development programming is not actually explicit about what the programme intends to do. What do feminist principles bring to discussions of "what working looks like?" and how do they inform the ways that Equity-focused evaluation's define and identify whether policies, programmes, projects, or organizations are successful or not?

Figuring out whether something works or not often entails first articulating what 'working' or success, means. For example, using Jandhalaya's (forthcoming) example on a women's education and empowerment programme in India, the donor defined success as 'consolidating and expanding' the programme to promote equality. Among a much longer set of goals, the 'engendered' goals developed by the programme included increasing 'life-long learning, leadership among poor and most marginalised women, and breaking discriminatory social barriers and practices (at individual, family, community and state levels).'

The two stakeholders set up success (and the measures of success) very differently[2]. Rather than counting increases in the number of groups and the numbers of women involved in groups, the engendered goal takes the programme down a path where success is defined by questions of 'which women' are involved, and how the groups make decisions, for example. The engendered framework sees the women's collectives as the place where success resides, where the programme is trying to bring change and thus where change should be measured. The debate over measures illustrates how evaluation can become a space where programmes are contested. They are contested in terms of how success is defined, where success is seen to reside, what is measured, and by extension often, what is done. The point here is not to suggest which definition of success is 'right' and which measures of success come closest to measuring changes that are relevant and meaningful. Rather it is demonstrating that in any programme, and certainly in large scale programmes, there are different or competing definitions and criteria of success. Equity-focused evaluation can bring those criteria to the surface for debate and critique.

Taking another example, in a women's health and empowerment project undertaken in the early days of the women's micro-credit movement, some of the changes that implementing organizations argued were the most impressive result of years of efforts in certain contexts, were that women from different castes were eating together, and that women were looking others (non-family members) in the eye (Bhirdikar et al, 2005). Other goals, perhaps much more impressive on paper, they were less proud of as they were much easier to achieve, for example, the thousands of groups that were formed. As Sudarshan and Sharma (forthcoming) point out, 'impact' is relative; measures should be embedded in the context of the intervention.

Evaluation practice

Inequities are deeply persistent for many reasons, many of which are deeply resistant and difficult to change. If reducing inequities is a goal of Equity-focused evaluation, it should come with a recognition of what can be at times a deep and inherent contradiction in using a time-bound, resource-bound, judgment-focused exercise – evaluation – with stakeholders who may be more or (often) less connected to equity and where the starting point is usually terms of

---

2   The full version of the original and engendered results framework is available in Jandyala (2010)

reference developed by others, to understand and shift inequities. Reflecting on this tension can lead to hand-wringing from evaluators deeply committed to Equity-focused evaluation (because it is never quite good enough, deep enough, transformative enough). Contributions on 'self-reflexivity' from feminist theory (Ackerly and True 2009, Cornwall 2003. Desai 2007) can offer important insights into Equity-focused evaluation on the role of reflection, and adjustment based on reflection.

Reflexivity is the critical introspection and analysis of the self as evaluator, and the way this influences the conduct of evaluation. Reflexivity can lead to insights and new hypotheses by pushing the evaluator to challenge their own theoretical positions. Reflexivity comes with a grounding or positioning of the evaluator within the process. The tension can become a healthy tension, because it comes with reflection, analysis, and adjustment, which lead to evaluation that moves closer to the Equity-focused evaluation principles. Sudarshan and Sharma (forthcoming), write: "It is our position that evaluation of NGO interventions in remote and difficult to work areas, has to maintain a balance between a level of objectivity and a level of sympathetic understanding." This positioning relates to questions of what is negotiable and non-negotiable in evaluation work sought, considered, or rejected, and to how evaluators see themselves in the evaluation process and, how this relates to the way they see themselves in the broader contexts in which they operate. It often finds a home in the way that evaluators identify themselves and position themselves in the work. Positioning is not new to social science. Feminist standpoint theory (Harding, 2004, Hartsock, 1997) has informed feminist research for over two decades and this understanding has filtered into the ways and approaches that some equity-focused evaluators use in their practice. 'Standpoint theory' offers alternative conceptualizations of rigour and validity rooted in principles of situated and constructed knowledge that acknowledges 'positionality', and begin from 'lived experiences.' Unlike 'positivist approaches', constructivist approaches' to evaluation start with the idea that knowledge is constructed and shaped, and the principle of evaluation should add value to those who are marginalized. As Harsh Mander (2010) has said of social science, Equity-focused evaluation: "is not an investigation into inert, static, external realities, but into the fluid, subjective worlds of people's lives, as experienced, interpreted, recalled and mediated by them." Similarly, Khanna (forthcoming) writes that providing "opportunities for the 'target community' to articulate their concerns and their analysis, and to share with them

my own analysis and suggestions, I believe promotes their empow-
erment."

Even if evaluators recognize that evaluation findings are socially
constructed and evaluators clearly shape that construction, calls for
'objectivity' can be used to discredit evaluation findings. This may
be particularly true in certain domains, and is likely to be particularly
true where evaluations challenge deeply entrenched inequities.
Equity-focused evaluations drawing on constructivist approaches
may find theoretical work such as 'standpoint theory' useful in artic-
ulating the rigour and validity of such approaches.

In contexts that privilege certain methods and approaches, Equity-
focused evaluators need language for demonstrating and speaking
to the strengths, rigour, validity (and limitations) of the approaches
they are using. In contexts where other evaluation approaches top
the lists of what donors and national governments consider cred-
ible, such work is both essential and contested.

Use of evaluation

A challenge in Equity-focused evaluation is the evaluator's responsi-
bility to promote use of or action on findings. While use is generally
never fully in the hands of the evaluator, Equity-focused evaluation
entails seeking pathways to use, while understanding these path-
ways as being political and negotiated. Certain pathways are risky
(for the programme, the group experiencing inequities, etc.), others
may be blocked, and some are strategic – but all are negotiated
and constructed. While Utilization-focused evaluation offers deep
insights and lessons for Equity-focused evaluation on designing for
use, feminist analysis also offers insights on the types of use that
are appropriate in Equity-focused evaluation. Put another way, with
an equity lens, 'any use' is not a valid or responsible use. Improv-
ing or addressing inequities must be a central use, along with other
uses particular to particular evaluations. Speaking to this idea of
responsible use, Sudarshan and Sharma (forthcoming) write:

> "Responsible feminism requires recognition of the contextual
> constraints and the feasibility of recommended courses of
> action and choices. We have therefore tried to be responsible
> feminists – pointing out specific changes and actions that in our
> analysis would empower women; at the same time, we have
> learnt a great deal about what is possible or desirable, given any
> particular context and capacities, and this learning itself moder-
> ates our recommendations."

Equity-focused evaluation should build responsible use explicitly into their evaluation design and process. This is mentioned earlier in this volume with various emphases being placed on ethics. In some contexts, pathways for Equity-focused evaluation findings to gain traction and influence are severely curtailed. For example, reflecting on one experience in India, Khanna (forthcoming) writes on the limits to use in the face of deeply internalized power differentials:

> "I found senior nurses and Nursing College Principals who were Master Trainers – and expected to be change agents within the profession – playing subservient hand maidens to Deans of Medical Colleges and State Health Officers. The same women, when interacting with their own junior colleagues, replicated exploitative relationships that they alleged they were victims of vis-à-vis the medical profession. The lack of self-awareness in relation to the concept of gender power relations and the lack of internal collectivisation in the face of external threats to the profession was very apparent. These issues could not be addressed within the evaluation debriefing. They needed a different, more personally introspective process…"

The underlying structures and systems that create inequities cannot be programmed away within contexts that perpetuate and reinforce those systems. Multiple pathways will generally need to be sought in Equity-focused evaluations, and used at different points in time (both immediate and longer term), including through policy and programming windows that open after the evaluation has ended.

Understanding how evidence informs policymaking and decision-making in different contexts is essential to commissioning and leading the Equity-focused evaluation's that will be used. For example, the Rajinder Sachar Committee Report (2006), on the state of Muslims in India, documented extensive and persistent inequities that combined to make the Muslim community among the most marginalized in the country. The report recommended autonomous evaluation of the extent to which programmes address issues of inequities. Media reports have quoted a former member secretary of the Sachar Committee, as saying that the authority met only three times in four years, did not consider implementing the report and had no independent technical person[3]. Though government-led evaluation of the implementation of the report did not occur, a research organization called the Centre for Equity Studies (CES)

---

3   http://www.hindustantimes.com/StoryPage/Print/728651.aspx

led by a member of the National Advisory Council, Harsh Mander did evaluate the implementation of the committee findings. Those researchers found that conditions were not improving, nor being adequately addressed or resourced, and blamed the government for lacking 'political courage' to directly address Muslims for fear of being criticised. According to news accounts, Minority Affairs Minister disputed the study, purportedly also arguing that it was constitutionally not possible to directly target Muslims in programmes and schemes. This example illustrates how discussion around inequities and targeting of inequities is often highly politicized. Evaluation of policies and programmes is not detached from those politics, particularly when it comes to use. In this case, Muslims face discrimination exactly because of their socio-religious identities, however, visible programme interventions directly targeting particular religious groups is deemed to be politically untenable. One way of side-stepping this issue is to critique the rigour of the evaluation studies and discredit the findings.

In terms of use by policymakers, certainly evidence suggests that their on-going involvement can be an effective strategy when they are open to evaluation findings in the area of enquiry. However, evaluators should not underestimate the degree to which findings that contradict the dominant policy discourse may find difficulty in getting traction with policymakers.

The work of Mercedes Gonzalez De la Rocha , on poverty in Mexico is an interesting example of how the policy context affects what findings are used. In the 1980s economic crisis De la Rocha's looked at poor people's strategies for survival and the ways in which poor urban households responded to crisis. Her work created the 'myth of survival' or the idea that the poor have an unlimited capacity to withstand shocks. About 10 years later, her later research following the 1998 financial crisis brought this 'myth' into question. She argues that her work has since then been selectively used by key institutions such as the World Bank, with her early work highlighting the strategies of the poor being picked up, and her later work, showing the limitations of those strategies, being ignored (2007). Evidence can, and often is, interpreted and used to reinforce dominant policies, in this case economic liberalization. Her first set of research supported existing economic liberalization policies – her later research did not. Evidence is usually used selectively. Those involved in Equity-focused evaluations need to be more strategic in understanding the role of evidence in policymaking and be more intentional in trying to support more open and transparent dialogue on evidence.

Engagement with and connections between social activists, researchers, and evaluators can also create or seize policy windows where there is openness to change, demand for change, and evidence to inform change. For example, Ramachandran (forthcoming) reflects on experiences from the education sector:

> "When the national assessment of gender and equity in primary education was presented to the government in a Joint Review Mission in 2002, the first reaction was dismissive – some said "there is no segregation in Indian education" and some officials objected to the use of the phrase "hierarchies of access". However, as the months rolled by and as commentators started comparing the findings with other research studies, especially the PROBE [Public Report on Basic Education] study, there was a gradual thawing. At least on paper, the government and donors accepted the findings and said that they would address it. The issue of children from different social strata attending differently endowed schools, gender discrimination in the choice of school by parents (government for girls and private for boys) or the issue of poorly endowed village schools / single teacher schools being the preserve of the most deprived – have now been accepted within educational discourse. The 2009 Right to Education Act has formally recognised the need to provide equal education for all."

Given the amount of evaluation happening there is a lost opportunity for synthesizing and generating deeper understandings from evaluation, on how development affects change on the ground. There is tremendous but largely untapped potential for evaluation to deepen understanding around inequities. Knowledge being generated through evaluations is generally not broadly shared, made available, or tapped and used to explore questions beyond the particular evaluation. This limits the opportunities to use evaluation to triangulate, challenge, or reinforce other bodies of knowledge around issues of social change and equity. Resources to generate evidence on inequities are always limited; groups working to address inequities in different domains and / or focused on different marginalized groups, need to begin to see evaluations as a rich body of evidence to drawn into other modes of knowledge generation and translation. Groups involved in doing Equity-focused evaluations need to begin to question decisions on keeping evaluations out of the public domain.

Conclusion

This chapter examined how principles drawing from feminist and other research traditions and theories can inform the framing, methods, and conduct of Equity-focused evaluation. This chapter has demonstrated how these principles can be used in practice to inform the understanding of programme theory, shape evaluation design and methods, negotiate judgment of success, guide practice, and guide choices and opportunities for influence. The chapter suggests that principles generated from feminist theory can be helpfully and usefully applied to strengthen Equity-focused evaluations – whether they have a central focus on gender inequities or on other inequities. Applying these principles to Equity-focused evaluation can help evaluation play a stronger role in understanding how societies change and which policies and programmes show promise in shifting norms and inequities. Reaching this potential requires more intentional integration of evaluation knowledge into broader knowledge translation exercises around rights, exclusions, and discrimination.

References

Ackerly, B. and True, J. (2009). Reflexivity in Practice; Power and Ethics in Feminist Research on International Relations. International Studies Review.

Birdikar, K. et al. (2005). How a Use and Learning Oriented Planning, Monitoring & Evaluation System Evolved in the Swayamsiddha Project. Available at: http://web.idrc.ca/uploads/user-S/11592100393M__E_Report_Version_21.pdf

Brewer, Rose M., Cecilia A. Conrad and Mary C. King, (eds), (2002). A special issue on gender, color, caste and class. Feminist Economics, 8 (2).

Budlender, D., D. Elson, G. Hewitt and T. Mukhopadhyay. (2002). *Gender Budgets Make Cents*. London: Commonwealth Secretariat Publications.

Çagatay, N., D. Elson and C. Grown, (eds), (1995). Special issue on gender, adjustment and macroeconomics, World Development 23 (11).

Chambers, R. (1983). *Rural Development – Putting the Last First*. Longman.

Chambers, R. (1997) *Whose Reality Counts: Putting the First Last*. Intermediate Technology.

Cornwall, A. (2003). Whose Voices? Whose Choices? Reflections on Gender and Participatory Development. *World Development* 31, 8: 1325-1342.

Desai, M. (2007). The Messy Relationship between Feminisms and Globalizations. *Gender Society* 21, 6: 797-803.

Devika, J. and Thampi, V. (2010). Empowerment or politicization? The limits of gender inclusiveness of Kerala's political decentralization. In, *Development, democracy, and the state critiquing the Kerala model of development*. Ed. K Ravi Raman. Routledge, New York.

Dietz, M.G. (2003). Current Controversies in Feminist Theory. *Annual Review of Political Science* 6:300-431.

Figart. D., E. Mutari and M. Power. (2002). *Living Wages, Equal Wages*. London: Routledge.

Harding, S. (2004). Rethinking Standpoint Epistemology. In S. Harding (Ed.), *The Feminist Standpoint Theory Reader*. New York and London: Routledge.

Grown, C., D. Elson and N. Çagatay, (eds), (2000). Special issue on growth, trade, finance and gender inequality. *World Development* 28(7).

Hartsock, N. (1983). The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism. In S. Harding and M.B. Hintikka (Eds.), *Discovering Reality: Feminist Perspectives in Epistemology, Metaphysics*.

Hay, K., and Sudarshan, R. (2010). Making Research Matter in South Asia. *Economic & Political Weekly, Vol xlv no 3*.

Hay, K. (2011). Evaluation Field Building in South Asia: Reflections, Anecdotes and Questions. *American Journal of Evaluation*. Vol. 31, Number 2.

Jandhyala, K. (forthcoming). *Indian Journal of Gender Studies.* Sage, India.

Khanna, R. (forthcoming). *Indian Journal of Gender Studies*. Sage, India.

MacKinnon, C.A. (2006). *Are Women Human? And Other International Dialogues*. Cambridge, MA: Belknap Press of Harvard University Press.

Mander, Harsh (2010). Words from the Heart: Researching People's Stories. *Journal of Human Rights Practice* (2010) 2(2): 252-270.

Mark, M.M., Henry, G.T., and Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco: Jossey-Bass.

McElroy, M. B. and Horney, M. J. (1981). Nash bargained household decisions. *International Economic Review* 22: 333-349.

Mertens, D. (2009). *Transformative Research and Evaluation*, New York; Guilford Press.

Mohanty, C.T. (2003). *Feminism without Borders: Decolonizing Theory, Practicing Solidarity*. Durham, NC: Due University Press.

Morell, J.A. (2005). Evaluation? Why Are There Unintended Consequences of Program Action, and What Are the Implications for Doing Evaluation? *American Journal of Evaluation*, 26-444.

Murthy, Ranjani. K. (forthcoming). Indian Journal of Gender Studies. Sage, India.

Narayan, U. (1997). *Dislocating Culture: Identities, Traditions, and Third World Feminism*. New York: Routledge.

Nash, Kate. (2002). Human Rights for Women: An Argument for 'Deconstructive Equality.' *Economy and Society* 31, 3:414-433.

Pawson, R. (2006). *Evidence-based Policy: A Realist Perspective.* London, Sage.

Patton, M.Q. (2010). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York: Guilford Press.

Podems, D. (2010). Feminist Evaluation and Gender Approaches: There's a Difference? *Journal of Multi-Disciplinary Evaluation*. Vol 6, No 14 (2010)

Ramachandran, V. (forthcoming). *Indian Journal of Gender Studies*. Sage, India.

Segone, M. and Bamberger, M. (2011). *How to design and manage Equity-focused evaluations.* UNICEF.

Sen, A. (1990). More than 100 million women are missing. *The New York Review of Books*, Dec. 20.

Sudarshan, R and Sharma, D. (forthcoming). *Indian Journal of Gender Studies Sage*, India.

Waring, M. (1988). *If Women Counted*. San Francisco: Harper & Row.

DECOLONIZING EVALUATION IN A DEVELOPING WORLD. IMPLICATIONS AND CAUTIONS FOR EQUITY-FOCUSED EVALUATION[1]

Rodney K. Hopson, Duquesne University;
Karen E. Kirkhart, Syracuse University;
Katrina L. Bledsoe, Education Development Center, Inc.

Introduction

This chapter suggests that the good intentions of Equity-focused Evaluation must be tempered by cautions. This concern flows from a legacy of research and evaluation that has exerted colonizing influences over Indigenous and minoritized populations. The opening section covers the context of development, evaluation, and culture. The second section argues that efforts to decolonize evaluation must begin with epistemology. A third section examines the implications of decolonization for evaluation method. Within the paper, a scenario is provided based on a development project in southern Africa. The scenario illustrates the complexity of stakeholders, projects, and cultural dynamics in a development evaluation where equity is an important concern. The chapter closes with implications and cautions for how evaluation generally, and more specifically, Equity-focused evaluation may perpetuate colonizing assumptions and aims.

Locating development, evaluation, and culture

Basic understandings of culture and cultural difference in the evaluation and development fields are only now taking shape; however, these have not been typically perceived as mainstream models of evaluation or development. The recent moves to address culture and cultural differences represent both a challenge and an opportunity for practitioners in development – specifically international

development – and for evaluation. There is a need to rethink dominant notions about how to work with and understand the particular concerns of traditionally marginalized and under-represented ethnic and racial groups, including indigenous groups, in varied country and nation-state contexts.[2] Now more than ever, both evaluators and developers are embracing specific paradigms that are more focused on social equity, empowerment, transformation, participation, and collaboration. Matters of culture and cultural differences are integral to these current approaches and deserve increasing attention.

For development practitioners and policymakers, the result is an increasing need to address questions and issues to do with culture and development. Thierry Verhelst's (1987/1990) book on this topic is just one example that critiques the collapse of current development models in favor of prioritizing a cultural dimension as an alternative path of development. Verhelst's premise rests on the failure of western development policies and models that have perceived nations in the global south as both backward and underdeveloped. This failure is despite the credible efforts by the United Nations in the 1980s to raise attention to these issues. Nevertheless, the prevailing development ideology, reinforced by notions of social Darwinism, colonialism, and unilinear views of history, has contributed to competing if not conflicting notions of development, modernization, civilization, and progress (Escobar, 1995; Rodney, 1981). These ideological conflicts persist between those from the West who practice development and those on the ground in "developing nations" who face "development" (Ferguson, 1994).

As such, even contemporary development practices, often referred to as participatory, "tend to emphasize the *who* and the *what* of development, with little attention to the *why* and the *how* [emphasis added]" (Eversole, 2005, p. 298). That is, development practitioners and policymakers ask questions such as who is involved in the development initiative and who benefits from its implementation. Other questions focus on illustrating distinctions between grassroots and top-down development and the different agendas that they imply. The 'what' development questions include, what initiatives are implemented; what do the development initiatives actually promote; or what type of development is to be carried out,

---

2   Of course issues of culture and cultural differences are much broader, such as around differences surrounding geography, gender, social class, etc... than around traditionally marginalized, stigmatized groups who are considered ethnic or/and racial minorities or indigenous, but the concerns of these groups are especially acute and the focus of this chapter.

"with the assumption that the right kind of initiative will solve those lurking questions about achieving authentic participation for target groups" (Eversole, 2005, p. 299).

'Why' and 'how' questions not only help to pay attention to cultural differences and the cultural 'groundedness' of efforts in development initiatives, but also dig deeper to draw on unstated assumptions about development, its processes, goals, and expected outcomes. These questions help to understand the competing and conflicting visions of development between the concerns and perspectives of different target groups, and outside development policymaking and practitioners. Asking questions about the 'why' and the 'how' of development presupposes a different set of interactions with a wide range of people, including acknowledging that involvement or participation by target groups without processes that fit the culture of the target group is hardly truly participatory.

Considerations of culture in evaluation have seen considerable growth domestically and internationally in recent years, since the first sets of literature on the topic in the mid-1980s and early 1990s (Hopson, 2003; Madison, 1992; Patton, 1985). What is clear from the early attention to culture is the bifurcated ways in which the evaluations and evaluators addressed the topic of culture. For those in the field working in international cross-cultural settings, the Patton (1985) edited volume asked, "what happens when we export the ideas, concepts, models, methods, and values of evaluators to other countries and cultures?" Evaluators from generally western perspectives were encouraged to consider the problems and potentials of doing international, cross-cultural evaluations. The Madison (1992) edited volume brought attention to how evaluations should be sensitive to multicultural issues and traditionally underrepresented and minoritized peoples and their perspectives in North American settings. Together, these volumes laid the groundwork for the increased number of professional workshops, conferences, symposia and published matter, on topics related to culture and evaluation in the United States (Chouinard & Cousins, 2009; Frierson, Hood, Hughes, & Thomas, 2010; Hood, 1998, 2001; Hood, Hopson, & Frierson, 2005; Mertens & Hopson, 2006; Samuels & Ryan, 2011).[3]

---

3   Much of this work has promoted an emerging evaluation approach or model, Culturally Responsive Evaluation (CRE), as a system or culmination of practical frameworks that attend to culture during various stages of the evaluation process.

Epistemological issues in decolonizing evaluation

Evaluation directed toward equity must comprehend and respect the cultural contexts in which the work is sited. This understanding goes beyond superficial appreciation of local culture as art, music, dance, and literature (Verhelst, 1987/1990) to deeper roots of history, spirituality, and core values. Central to the cultural location of any inquiry – research or evaluation – are cultural assumptions and beliefs about the nature of knowledge itself. Thus, any evaluation that hopes to address equity must begin by considering epistemology.

Epistemology is foundational to all evaluation, but it plays a particularly key role in those methodologies that seek to promote equity and social justice. Epistemology is both personal and political. Evaluators must reflect on what knowledge they privilege as well as acknowledging the politics of knowledge construction. Examining the political nature of knowledge construction reveals the influence white privilege in marginalizing alternative ways of knowing, which leaves Indigenous inquiry "off the 'buffet table' of methodological options" (Kovach, 2010, p. 79). Decolonizing epistemology broadens and expands what knowledge can entail, and creates room at the table for a richer menu of options.

The colonizing effect of research is now well recognized (Estrada, 2005; Grande, 2008; Smith, 1999; Weber-Pillwax, 1999). Indigenous frameworks that are explicitly decolonizing have been proposed in both research (Cajete, 2000; Estrada, 2005; Kovach, 2010; Smith, 1999; Wilson, 2008) and evaluation (Kawakami, Aton, Cram, Lai, & Porima, 2008; LaFrance & Nichols, 2010). In evaluation, colonization manifests itself in determinations of merit or worth that are defined from a non-Indigenous – often geographically and culturally distant – perspective, and applied to Indigenous persons and programmes without regard to local culture and values. Evaluation is colonizing when it defines the programme, poses evaluation questions, gathers and analyzes data, and formulates results in ways that silence the voices of Indigenous persons in the local context. Evaluation is colonizing when, by omission or commission, it bolsters majority power structures without critique or challenge. Decolonizing evaluation means locating it within Indigenous cultural specificities, preferences and practices. It means recognizing and critically interrogating Eurocentric knowledge systems and standards of inquiry that have historically been imposed upon Indigenous cultures in Africa, Asia and the Americas.

The key to shifting perspectives lies in understanding how knowledge is created and understood within a given cultural context and in using that understanding to define standards of good practice in research or evaluation, including standards of legitimation such as validity. Standards of practice and legitimation in turn guide method choice. These three elements – standards, validity and methods – stand in reciprocal relation to one another, grounded in epistemology. Together, they offer a three-strand approach to decolonizing evaluation, informed by Indigenous scholars on five continents.

Strand 1: Epistemology informs what is understood as good evaluation practice. To decolonize ways of thinking about good evaluation practice, one must appreciate how Indigenous epistemology places the entire evaluation process outside of Western[4] understandings. Indigenous epistemologies stand in contrast to Western epistemologies in both scope and content. The boundaries of what constitutes knowledge are much broader; knowledge from dreams, visions and prayers is respected (Ermine, 1999). Historically, Indigenous knowledge systems, which do not separate spirit and reason (Deloria, 1999), were dismissed as superstitions by Eurocentric researchers. "Indigenous knowledges could not be understood from a reductionist analysis because they could not be fragmented, externalized, and objectified" (Kovach, 2010, p. 77). Little Bear (2000) reviews fundamental differences in Aboriginal and Eurocentric worldviews. He characterizes Aboriginal worldviews as holistic and cyclical, generalist, process-oriented and grounded in place; European worldviews as linear, singular, static and objective. It is important to note that when he speaks of "jagged worldviews colliding," Little Bear is speaking not only of collisions between/among cultural groups but also within a given person.

To decolonize means that standards of good evaluation practice are defined by local values and protocols. Inquiry must be vetted through appropriate local authority that determines what knowledge can be shared under what circumstances. Evaluation must be done in the right way, following the path recognized as good, honorable, and respectful – all of which is defined by local culture, linked to understandings of the right way to live upon the earth.

Indigenous belief in the interconnectedness of all living things disrupts linear understandings, both in defining the parameters of what

---

4   We use Western here and throughout to designate those perspectives that generally derive from a limited North American (United States, primarily) and western European geographic reality that privilege theirs above others.

is being evaluated and in setting forth the framework for the evaluation. Telling a programme's story replaces linear Western logic models (LaFrance & Nichols, 2009). Cram (2011) cites Conner's 'Draw-a-Path' process of creating a visual timeline with drawings and symbols so that people can tell the story of the programme as they understand it. Portrayals of Indigenous evaluation frameworks may also be visual and non-linear, often employing story or metaphor. They are in sharp contrast to Eurocentric portrayals of approaches to evaluation practice. Where Western frameworks or models are often presented via steps or categories, Indigenous frameworks use visual images and extended metaphors. For example, Estrada (2005) uses the metaphor of the *Ceiba* (Tree of Life), to connect research to a circular and multidimensional cosmology directed by the Maya sacred book of Creation, the *Popul Vuh*. LaFrance and Nichols (2009) use Dr. Eric Jolly's story of Cherokee basket making, told by his grandmother, as a metaphor for Indigenous evaluation.

Good evaluation under Indigenous epistemology addresses power imbalance. It positions evaluation to resist exploitation and oppression by centering control of the initiation, information-gathering procedures, interpretation, and sharing of information in the local community (Bishop, 1998, L. T. Smith, 1999; G. H. Smith, 2004). Evaluators must respect local authority and protocols for entering a community and making introductions. These protocols are culturally-specific. For example, Bishop (1998), a Maori scholar, describes a formal ritualized introduction, *mihimihi*, as "a statement of where you are from and of how you can be related to these other people and the land, in both the past and the present" (p. 203). LaFrance and Nichols (2009) introduce themselves through tribal and clan affiliation, family genealogy, and geographic locations in which they are grounded.

When interacting with persons of authority, such as elders, to gather information, evaluators must consciously relinquish control of the conversation and let the story unfold from the perspective of the teller. The conversation should not be shaped to fit a predetermined outline or interview schedule without ensuring the questions or items are sufficiently rooted in Indigenous realities. Power dynamics relate also to the underlying vision of evaluation bringing benefit to communities. Who gets to define benefit and what segments of a community or society will experience it? These questions relate directly to concerns for equity.

Location is extremely important to the practice of Indigenous evaluation. Both the evaluation and the evaluator must be situated in

context. Place is a living presence that defines nationhood and the core values of Indigenous peoples (LaFrance & Nichols, 2009). So too, knowledge and the ability to access it may be linked to a particular place or location (Hermes, 1998; Kovach, 2010; Wilson, 2008). Location also refers to the cultural location of the evaluators and their use of theory as well as to the programmes, structures or persons being evaluated (Kirkhart, 2010). For example, in her *Ceiba* metaphor, Estrada (2005) sees the tree trunk as Indigenous knowledges and woman-centered epistemologies, noting that "this stance better reflects my cultural location, since the Maya culture is traditionally matrilineal" (p. 50).

Time plays an enormous role in setting the parameters of good evaluation. Indigenous epistemology calls evaluation to look forward and backward, beyond the present moment. Long-term outcomes must be carefully considered. Evaluators must resist becoming narrowly focused on immediate effects on direct programme participants; impacts (intended or unintended) on persons in the community are of equal or greater importance. Indigenous evaluation is patient. It may not conform well to fixed deadlines or the timelines of Western funding cycles. Standards of practice also address time in a way that is unfamiliar to non-Indigenous evaluators. "Good practice" includes rules governing *when* certain stories may be told or information gathered, often tied to the four seasons or cycles of Nature (Tafoya, 1995).

To achieve a decolonizing outcome, Indigenous evaluation supports sovereignty and self-determination. Local epistemology gives Indigenous persons control over how evaluation is conducted, in ways that contribute to greater control over their lives. The conduct of Indigenous evaluation honors core values (LaFrance & Nichols, 2009) and reflects cultural, political, economic and social preferences (Smith, 2004). It is action oriented. Indigenous evaluation practice moves beyond determinations of merit or worth to support cultural reclamation and transformation.

Indigenous epistemology expands voice and vision in evaluation. It offers more diverse worldviews (and a correspondingly clearer sense of the limitations of narrow cultural location) that challenge our prior understandings, moving evaluation beyond a data-gathering and interpretation exercise, even within participatory or collaborative models. A similar expansion recasts validity concerns.

Strand 2: Epistemology informs validity. Validity holds authority in systems of inquiry – both research and evaluation. It signi-

fies power and control over the legitimization and representation of knowledge (Bishop, 1998), which is contested space in decolonization. Who determines what is valid and invalid, legitimate and illegitimate? What is given heavy consideration and what is discounted? Under decolonization, validity must be expanded beyond a reductive, monolithic construct to allow different ways of being valid and multiple pathways of validation and legitimization. Validity must be recast to fit local understandings about the nature of knowledge and how legitimacy and trustworthiness is determined.

Validity resides within language; this is especially critical for Indigenous knowledge and languages. Translation compromises and disrupts valid understanding. If one is limited by the English language for instance, genuine meaning may not be conveyed. Nuances of meaning are not captured in generic nouns. A simplistic search for direct translation leads to frustration and misunderstanding (see Tafoya, 1995, for an illustrative story).

Decolonization involves unfolding validity arguments to make the logic and bases of justification more transparent and to expand the range of arguments considered legitimate. Indigenous understandings of validity emphasize justifications that are relational, experiential, and attentive to consequences (Kirkhart, 1995, 2005). Relational accountability lies at the heart of Indigenous research and evaluation (Wilson, 2008). Relational criteria replace criteria of "neutrality, objectivity and distance," which have historically excluded Indigenous peoples from participating in the construction, validation, and legitimization of knowledge (Bishop, 1998, p. 201). Experience resides in both outward and inward space, physical and metaphysical, objective and subjective (Ermine, 1999). Consequences are viewed in terms of the good of the whole – the sovereignty and well-being of the tribe or community (LaFrance & Nichols, 2010).

Decolonization avoids the construction of sweeping rules and procedures for validation that lie outside a given research or evaluation project, because such rules remove authority from participants. When validity is located external to context, it ends up defining other people's realities, disregarding local epistemology (Bishop, 1998). Validity must be understood within, not outside of, Indigenous worldviews.

Indigenous perspectives on validity must be foremost, so that conclusions are validated by appropriate criteria and procedures compatible with the context. But this does not mean that all Indigenous

perspectives are identical or that *only* indigenous perspectives may apply. The links between global and local must be well understood and made visible in evaluation research (Stanfield, 2011). Such linkage is an issue of particular importance to matters of equity that cross national boundaries. Evaluators working in international development contexts import paradigms that do not fit the realities of the local cultural context, including but not limited to frameworks of validation. These imported beliefs must be recognized, understood, and replaced or balanced with culturally-specific ones. The importance of balancing rather than replacing is not an unwillingness to commit to local epistemology. Rather, it is recognition that the primary value in evaluation is utility – in this case, the utility of a given approach in achieving equity. One must be multilingual with respect to epistemologies, because conversations about access to resources cross cultural boundaries (Williams, 2006). Wholesale rejection of majority epistemologies may be as unproductive as wholesale acceptance. In a development context, the ability to move clearly and transparently among epistemologies in working with multiple stakeholders, avoids "collisions" of incompatible worldviews in favor of well-informed dialogue.

Strand 3: Epistemology informs method. Consistent with broader understandings of good evaluation practice and validation, epistemology also expands the range of available methods. It alters time frames and modifies procedures for gathering, processing and using information to reflect local ways of knowing. This is not simply a question of "sensitivity to context" in translating and using non-Indigenous tools, but of the nature of the tools themselves. Tools used in Western approaches may need to be deconstructed and rethought. Some methods are culturally-specific and must be created from scratch. Others may be repurposed, modified or adapted to context. Tools that are traditional to Indigenous communities may appear "new" when viewed from Western perspectives. Conversely, methods that may at first glance appear familiar (e.g., interview or observation as data-gathering strategies) are in fact quite different when applied with a different epistemological foundation.

The processes of data gathering must allow participants to share their experiences in their own terms. The non-linearity of Indigenous knowledge discussed above (Little Bear, 2000) often defines a data-gathering process that does not resemble a Western exchange of questions and answers. In framing evaluations, questions are sometimes not appropriate at all; statements about what knowl-

edge is desired or needed are more suitable (LaFrance & Nichols, 2009). Answers often reveal themselves in stories that are indirect and non-linear. This requires a different attitude and skills set on the part of the listener, as well as those interpreting the stories to form evaluative conclusions.

> "It helps if you listen in circles, because there are stories inside and between stories, and finding your way through them is as easy and as hard as finding your way home. Part of finding is getting lost, and when you're lost you start to open up and listen" (Tafoya, 1995, pp. 11-12).

Along with tools and methods, Western notions of research design may also be reinvented. Preordinate designs, perhaps mapped out even prior to entering a community, are antithetical to Indigenous evaluation frameworks. Indigenous epistemology informs designs that are emergent. Hermes (1998) is explicit in stating that her methods were not clearly delineated before she started her work. They were continually changing, acting as a "situated response" to both theory and context. Such a fluid view often accompanies a focus on inductive (vs. deductive) methods.

Indigenous epistemology privileges context over method (LaFrance, Nichols, & Kirkhart, forthcoming). "Context-stripping methods" long associated with positivist epistemology (see, for example, Mishler, 1979) must be avoided in favor of methods that honor the context dependence of ideas ("variables") and the relationships among them. However, it is important to be clear that no single method or category of methods (e.g., qualitative vs. quantitative) necessarily obviates the problem of colonization (Bishop, 1998).

The scenario that follows illustrates how understanding context sets the stage for development evaluation. It describes the project history, key stakeholders, and other key aspects of an educational evaluation of the San[5] in the context of southern Africa.

San education evaluation project scenario: Part 1

In the southern African region of the world lives one of our oldest known human groups, the San. In one such community, an international aid agency based in Europe has requested its first evaluation of the last decade. Project activities included teacher training, teacher support, and other efforts to support education at a primary level and to influence the educational situation in the local

---

5   Other terms used to refer to San are Bushmen, or (in Botswana), Basarwa.

San community. At stake are a set of decisions and recommendations needed to contribute to administrative transitions underway by the Ministry of Education (MoE) and the European aid agency (EAA) regarding the future directions of the educational projects. At the time of the evaluation, there is a little angst about the future direction of the educational project from the ministry and the community. It is believed that the evaluation will help to answer questions and propose a future direction for the government and dispel the fears the community has about the potential termination of the project. In addition to the angst about the future of the project, there is as much anxiety about the evaluation. Some key stakeholders recall how a prior evaluation by another international aid agency over a decade ago left more questions than answers.

Since the development of language and educational projects after Independence, the San community has had less participation in formal schooling systems than any other indigenous group, both in the country and in the region. For the 20 years after Independence, efforts were made by other international and local aid agencies in collaboration with the MoE to increase the number of San attending school and to take advantage of the fruits of development. Although the San gained increased access to formal schooling during this time, there were incredible challenges for them in completing that schooling. For instance, while increases had been seen in the attendance of San children in formal schools, too many dropped out due to complex economic, social, and cultural issues that prevented many from achieving educational success. Specifically, reports of bullying from other ethnic indigenous groups, lack of supportive adults (teachers and staff), irrelevant curricula, and acute homesickness were the common reasons reported by nongovernmental and governmental agencies for why San children dropped out of schools at the upper primary and junior secondary levels. Typically, no San children remain enrolled at the senior secondary level of schooling.

Even before arriving at primary school, the San children have foundations in language(s), culture(s), values, and skills learned from their home and local communities. The village communities from which most rural San children come are tight-knit systems of parents and community leaders. Historically, the San lived as hunter-gatherers in areas of their traditional territory. At Independence, traditional ways of living continued to influence a series of important decisions and policies on maintaining and classifying traditional lands. While these decisions seemed to benefit this San commu-

nity, there were significant racial and class divisions and many San families at and following Independence still struggle with a basic subsistence livelihood as farm laborers, either on the outskirts of town or in rural areas.

As required by the terms of reference (ToR), the evaluation methodology was expected to be a review of project documentation through a desk study and interviews with key stakeholders involved in the project. More specifically, the ToR proposed: i) a general update on the situation for the San; ii) an overview of goals achieved compared to goals initially set forth; iii) analysis of effects of the project on grade 1-3 students, compared to those not participating in the project; iv) analysis of educational levels achieved by San children; v) an investigation of how partners perceive the success of the project operation, implementation, and influence; vi) input toward local ownership and local participation; vii) discussion of how funds were prioritized; and viii) recommendations for future steps. According to the ToR, the report of the evaluation would span a period of four months and a summary would be provided in the local language.

As a starting point, a team of foreign-born evaluators wrestled with the specific ToR, wondering to what extent the desk study and interviews would reveal answers to the questions posed; how an evaluation study could be designed to answer the questions related to achieved and intended goals; the effects of the project on students in primary school; and the expected influence of the project and the evaluation for future decision-making purposes. The EAA's managing director, in response to questions about the ToR, was both vague and brief. He responded to the evaluators' request for clarification by reiterating his hope that the evaluation would assist communities in focusing on the intended directions of the project. Of particular concern was the impending transfer of responsibilities for the project administration from the primarily European-led project coordinators, who had led the efforts for the first eight years, to the regional office of the MoE. The motives for the transfer from the international aid agency to the country-led regional line ministry office were not clear. The EAA had presumed that the time taken to develop the educational projects was sufficient to allow for the transfer, although at the time of the evaluation, considerable unease remained within the San community over the timing and specifics of the transfer.

Methodological considerations in decolonizing evaluation

The dilemma of the evaluators within the San scenario is one that all evaluators must face: evolving situations subject to whatever changes occur. Ensuring evaluation that promotes equity and social justice means understanding that methods and methodological designs must be representative of the culture, the location in which the culture is situated, and the political context.

Arguably, the tension between the consideration of Indigenous epistemology versus the practical constraints of resources, politics, and decision-making timelines is palpable (Kovach, 2010). To simply insinuate that conducting evaluation in developing countries is a challenge is to understate the issue of undertaking evaluation in unforeseen and, in many ways, unstable circumstances including culture, socioeconomic status, physical environment, and internal and regional politics. When one adds the agendas of westernized often Eurocentric countries such as the United States, that often fund developing countries' humanitarian programmes, the multitude of stakeholders who impose perspectives on what is considered good evaluation, grows exponentially.

Although only the most altruistic rationales underlie development programming and by extension, programme evaluation, development remains largely an externally focused activity, with funders and donors as the givers, and target groups as the receivers (Eversole, 2005). By its very nature, evaluation that is equity-focused or socially just should challenge the view of obtaining answers in a context-free location, and in a value-free climate. Instead, evaluation models that privilege equity, for instance, would strive to represent and explore the views of "the other," and dare say, work to decolonize by reflecting on the common core values that underlie the cultural context, the culture, and the political milieu.

Yet, there is pressure to answer the questions of developing countries (and funders) using a westernized, and arguably Eurocentric view of what is considered rigorous, systematic and objective evaluation. It is problematic to interpret data using a one-size-fits-all approach to answer questions from multiple stakeholder perspectives. Methods and methodological designs must be contemplative, exemplifying the context and various mores imbued throughout the country. In this way, the focus is on the values of the country and target group rather than a colonized view. Additionally, tailored

designs and methods are, in the end, intended to insure the precision, validity, and credibility of the information gathered.

The above mentioned issues have been considered as tensions in epistemology and practice in developing countries. For instance, in understanding evaluation in Africa, evaluators such as Gariba (2007) call for understanding African Knowledge Systems as a basis for evaluation, rather than using traditional perspectives. Such a perspective means considering: a) the historical and political perspectives of the country; b) the purpose of evaluation in the context and within the setting; c) what standards should be used to provide the background for design and execution in methodology and data collection; and d) what is meant by 'credibility of information'.

Understanding historical perspectives and contexts of countries and their relationship to evaluation methodology. To understand the context of the country and modern day programming, one must understand international relations in general. Pronk (2009) notes that international conflicts in the 20th century were, at a larger level, both political and ideological between west and east, but also reflective of the aspirations of people to free themselves from political, economic and cultural oppression. The 21st century has brought similar perspectives, with nations across the world witnessing a wave of resistance to colonization that has been imposed either within the country or with influence from powerful nations from east and west (Pronk, 2009). With the aforementioned as a backdrop, the focus on equity across nations, within classes and cultures, becomes significant.

Acknowledging imperialism in programming and evaluation. Conlin and Stirrat (2006) acknowledge that programmes in developing countries were and have continued to be primarily donor focused, designed and shaped in the manner that donors find most palatable, and most congruent with their values. By extension, evaluations face similar issues. Evaluations have in the past, and in many ways continue, to abide by the scientific clinical trial model of testing that includes Randomized Control Trials (RCT) and experimental designs. Despite calls for considerations of the contrary (e.g., Mertens & Wilson, 2012; Hood, Hopson, & Frierson, 2005), these designs are still considered a "gold standard" for impact evaluation, and the standard by which all evaluation, no matter what the culture, political perspective, or context, should aspire. Yet to Conlin and Stirrat (2006), this so-called gold standard is feasible only in a modicum of international circumstances and contexts. In under-

standing this reality, it becomes clear that evaluation that privileges equity must consider the context in which the evaluation would be conducted, and acknowledge the colonialist perspective that such a standard projects.

In attempting to address equity in evaluation, the design must address variables such as the historical understanding of imperialism. Such designs would account for issues such as class and status warfare, and why these might influence any type of international development that might be accomplished. For instance Gariba (2007) pondered the question of how African evaluators cope with challenges posed by the need to use evaluation as a tool for transformation, and not just ex-post assessment. Thus, evaluations committed to equity and social justice must keep in mind that the evaluative process is a collaborative process, a synergy between all parties dedicated to obtaining evidence that is representative of context, while addressing the needs of stakeholders.

Standards for design and execution in methodology and data collection. Many approaches – even those focused on social justice, inclusivity, and equity – are steeped in traditional deductive methods versus those methods that are more inductive. These deductive strategies may be useful and in some cases necessitated by the design; however, they can be disarming and distancing, discouraging the very aim of social justice values. Therefore, the focus must be on those designs that encourage inductive reasoning and by extension, designs. Conlin and Stirrat (2006) note that, "perhaps the time has come to recognize that interpretative approaches which owe more to history than to experimental sciences might be better suited to the world of development" (p. 200). Designs that are inclusive of non-traditional, inductive-based methods might be better suited to chronicling a community's response to programmes. Additionally, such methods can insure validity of information, as well as informing multicultural validity (Kirkhart, 2005).

In addressing what else might enhance an equity-focused agenda, the Culturally Responsive Evaluation (CRE) framework is informative in providing insights and strategies for promoting equity (Frierson, et al., 2010; Mertens, 2008) and is adept at acknowledging and encouraging the full participation of underserved and disadvantaged communities. CRE strives to bring the cultural context to the fore in evaluation design, thereby insuring that the methods used, whether traditional or non-traditional will be true to the context in which the culture operates.

Even under the best circumstances within a CRE approach, donors and recipients can, and regularly do, clash on factors related to programme design and by extension, the best way to illustrate results. As Gariba (2007) notes, the question is, whose standards for research are used? And, by association, whose ethics?

Cultural competence and responsiveness in method design and methodology. A mainstay for Equity-focused evaluation must be the consideration of cultural competence in method design and methodology. Cultural competence can be defined as a "state of being" (American Evaluation Association, 2011). But Lee and Farrell (2006) wonder if is cultural competence an excuse to continue to perpetuate racism, nationalist attitudes and stereotypes (for instance) in programming, and evaluation? They warn that the tendency to use race, culture, and ethnicity interchangeably further categorizes communities and groups. To avoid these issues, one must consider that there must be some sort of collaboration between those who are evaluated, the evaluation agenda, and the evaluator. As Gariba (2007) notes, those with power and instruments of communication and those who do not have access to these instruments (the target population) must work in tandem with one another, deciding the most congruous channels for communication of results.

Participants and participation. Much like development programming, evaluation donors of developed countries, while asking for a more collaborative and close relationship with developing countries, continue to subtly and sometimes overtly shape the goals, landscape and processes of initiatives. This colonial perspective suggests that there is a specific way to design programmes, usually one that is closer and more representative of a westernized perspective. In an Equity-focused evaluation, it makes sense that the values of those who are most affected by the programme and evaluation (the consumers) would come to the fore, while acknowledging the context in which these values are executed, and reconciling those of the funders and donors who may subscribe to a different perspective.

Credibility of information. Perhaps the cornerstone of methodology endeavors is the credibility of information gathered (and interpreted). Credibility, as Donaldson (2009) notes is the reliability and *validity* of the information gathered, the analysis of it in a manner that is reflective of the data, and finally, dissemination at all levels. Earlier within this chapter, we postulated that validity must fit local, contextual understandings about the nature of knowledge. Specifically, to insure true validity, multiple ways of knowing and concep-

tualizing must be encouraged and accepted. Thus, the evaluator is emboldened to look beyond traditional identifications and definitions of validity. These might include, as Gariba (2007) encourages, considerations of "voices and collective energies of the marginalized" (p. 8).

Although we do not assume that there are exact similarities between evaluation performed in developing counties versus developed countries, the ongoing conversation is appropriate for both contexts. Such context-specific situations illustrate the need for uniquely developed location-situated understandings and evaluations. For instance, the American Evaluation Association's Public Statement on Cultural Competence (2011) provides several tenets for culturally responsive and culturally competent evaluation, especially as they relate to validity. Validity requires a level of shared understanding, and to achieve such an understanding, all voices and perspectives must be honestly, equitably, and equally represented. Such a combination ensures that credibility of information is enhanced. When translating the tenets of validity for Equity-focused evaluation, into international development efforts, validity and by extension, credibility is supported and enhanced (AEA, 2011, pg. 5) when evaluators:

- accurately and respectfully reflect the life experiences and perspectives of programme participants in their evaluations;

- establish relationships that support trustworthy communication among all participants in the evaluation process;

- draw upon culturally relevant, and in some cases culturally specific, theory in the design of the evaluation and the interpretation of findings;

- select and implement design options and measurement strategies in ways that are compatible with the cultural context of the study;

- consider intended and unintended social consequences in the overall assessment of their work.

The final section of the educational evaluation scenario in the San community in southern Africa specifically addresses methodological considerations and challenges in carrying out equity and socially just evaluations. Specifically, the evaluators were placed in a conundrum in which they needed to choose a design that was responsive to the development agendas of the ToR, but also responsive to the needs of the San. In keeping with Eversole's (2005) participatory typology, the evaluators discovered that a collabora-

tive approach was best suited for the context, as well as the political landscape. However, as the scenario progresses, we notice that the collaborative approach does not solve all questions, and issues linger at the conclusion at the end of the scenario, as discussed below:

San education evaluation project scenario: Part 2

The evaluators chose a design that would be collaboratively formative and culturally grounded, ensuring there was opportunity to meet a host of stakeholders at multiple levels in the country, region, and community, including building an opportunity to share the draft report to all groups prior to submission to the EAA. They had had recent experience where researchers or evaluators had either collected information without full ownership of the document by the San community or, as was more typical, the findings were non-representative, not available, or not in a language that could be understood by most in the local community.

To illustrate the challenge that the evaluators faced at the time of their fieldwork, the total number of learners in grades 4-8 at the regional primary and junior secondary schools numbered only 40 out of the nearly 150 primary school students. These figures are for the village schools in a good year, based upon enrollment statistics recorded by the principal of the village schools. Many San children came from their village schools at grades 1-3. Out of the large numbers of students who had completed grade 3 over the last decade or so, many did not make the transition to school in the regional center. By the time of the evaluator's visit, there was only one student still enrolled in 8th grade and fewer than a handful in junior secondary school. As the MoE statistics continued to illustrate, the higher the grade, the more likely San learners were to drop out. The situation facing the one village school student left in 8th grade reflects a larger, more complicated challenge related to changes and transitions that took place at multiple school levels. These challenges were both pupil-related and staff-related. They included logistical challenges around transportation, materials, food, support, the increased need for teacher uptake and training, and community consultation and involvement.

In two in-country fieldwork periods and one set of interviews at the aid agency offices in Europe, a variety of stakeholders were interviewed at the locations where stakeholders worked. At the time of the interviews, these stakeholders were involved with educational efforts in the San community in the capital city, the regional

city, or the local or village constituency, as well as in the European country that funded the project. At the outset of the fieldwork, it was important for the evaluators to visit each village and ensure a community meeting was held to gather insights at the village level. Local research assistants were employed during the fieldwork at the regional and local levels. The nearly 130 respondents included educational planners, land and environment consultants, members of the MoE in regional positions, San community, principals, teachers, village committees, students, and matrons.

During a second fieldwork visit a few months after the initial team fieldwork, preliminary results were reported to groups of stakeholders in three regions of the country – in the capital city, in the regional headquarters, and in the San regional area. In the San regional area, two presentations were given over a weekend to facilitate attendance by groups of teachers and community members to listen and provide feedback. In addition to presenting the draft report, the purpose of the meetings was to gather additional input regarding the intended recommendations of the report. At these meetings, evaluators ensured that multiple languages – local, regional, and official – were spoken in the feedback presentations and that the final report recommendations would be in the San language.

In the final report to the EAA, evaluation recommendations were arranged in several sections related to the key sections of the report: the context and background of the aid agency's project, the goals of the project, its impact, and challenges for transition, logistics, roles and responsibilities, teacher training, and community consultation and involvement. Five priority recommendations preceded the remaining 49 recommendations. These priority recommendations focused on the key presence of the agency and the need for a consultative conference, a clearly established mission statement, role clarification between the aid agency and the ministry, and the development of alternative approaches to education.

The response to the report from the aid agency was generally positive and priority recommendations were, by and large, likely to be acted upon beginning with the suggested consultative conference later in the year. This was to involve a diverse group of stakeholders who would revisit the recommendations of the evaluation report. What was less clear from the perspective of the evaluators, despite the expected consultative conference where all matters pertaining to San education would be discussed, were the continued expressions of alienation these groups faced. This alienation was evident

even in government schools that were attempting to close achieve-ment gaps and to ensure that all groups had access to schools and educational materials that had heretofore not been available for most and especially the San. The nature and transparency of the relationship and roles between the MoE and the aid agency also remained unclear, as was how community involvement and input would actually be guaranteed or promoted in moving forward.

Concluding thoughts and critiques

Much of this chapter has been focused on understanding the impli-cations and cautions about Equity-focused evaluation within the context of westernized colonialistic programming, while acknowl-edging, embracing, and privileging local culture. As with all evalu-ation approaches, but especially those that are transformative and focus on social justice and Indigenous populations, the challenge is to provide voice and true representation of those who the evalu-ative process affects the most. We do not question outright the motives of developed countries such as the United States who would work collaboratively with developing nations. Many of those collaborations are altruistic in nature. However, we consider Equity-focused evaluation to be a strategy by which to *equalize* the often benevolent and charitable nature of international funding and pro-grammatic assistance. Given that Equity-focused evaluation is new to, as Mertens and Wilson (2012) would assert, values-driven eval-uation approaches, it bears a large responsibility.

Before closing this critique, transparency requires acknowledg-ing the complexities and pitfalls of what has been proposed, just as the scenario above insinuates. Decolonization is neither easily nor quickly accomplished; it is an enormous project over genera-tions of evaluators. It requires a long view of time, and an appre-ciation of the slow process of building trustworthy relationships. It requires deep commitment and motivation, as well as resilience to steel oneself against the realities and limitations of what can be accomplished at any one site and time – what Maori scholar and educator Graham Hingangaroa Smith calls "the politics of truth" in the context of his work in the Academy. It requires an apprecia-tion of irony, acknowledging that one is undertaking decolonization within a development context that has held, and often continues to hold, an explicitly colonial agenda. It requires an insider perspective that must be engaged authentically. When evaluators are not Indig-enous to a given community, it requires genuine partnerships and

an acute awareness of one's cultural location. It requires unlearning some traditional definitions of "good evaluation" to create space for broader understandings of validity and rigor, of utility and relevance. It requires tolerance for ambiguity, moving away from a search for singular truths.

This chapter suggests that decolonizing evaluation is not easy work. There are many challenges in committing to and carrying out Equity-focused evaluation. There are deep commitments to the epistemologies and methods in which we have been schooled. Furthermore, there is arrogance in assuming that anyone can fully know and understand the worldviews of others, particularly when one is coming from outside the community in which the evaluation is located. In grappling with new realities, one risks reifying cultural dimensions as singular and static or creating fixed categories to aid description. One must struggle to retain a vision of culture as fluid, dynamic, and multifaceted. In seeking to build evaluation on Indigenous knowledge, one risks advancing an assimilation agenda that changes the worldview one is attempting to incorporate. Seeking partnerships brings stressors of multiple accountabilities, creating tensions between two or more loyalties, and potentially dividing communities. One must balance circumspection about the possibilities of missteps with the imperative to move ahead toward equity and justice, not to be paralyzed by one's acknowledged ignorance but to advance with humility.

References

American Evaluation Association. (2011). *American Evaluation Association public statement on cultural competence in evaluation.* Fairhaven, MA: Author. Retrieved from www.eval.org/ccstatement.asp.

Bishop, R. (1998). Freeing ourselves from neo-colonial domination in research: A Maori approach to creating knowledge. *International Journal of Qualitative Studies in Education*, 11(2), 199-219.

Cajete, G. (2000). *Native science: Natural laws of interdependence.* Santa Fe, New Mexico: Clear Light Publishers.

Chouinard J., & Cousins, J. B. (2009). *A* review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation.* 30(4), 457-494.

Conlin, S., & Stirrat, R. L. (2008). Current challenges in development evaluation. *Evaluation, 14*(2), 193-208.

Cram, F. (2011, November). Methods for Kaupapa Maori Evaluation. Panel presentation at the annual meeting of the American Evaluation Association, Anaheim, CA.

Deloria, V., Jr. (1999). Ethnoscience and Indian realities. In B. Deloria, K. Foehner,
& S. Scinta (Eds.) *Spirit & reason: The Vine Deloria, Jr., reader* (pp. 63-71). Golden, CO:
Fulcrum Publishing.

Donaldson, S., (2009). In search of the blueprint for an evidence-based global society.
In S. Donaldson, C. Christie, C., & M. Mark (Eds.), *What counts as credible evidence
in applied research and evaluation practice* (pp. 2-18). Thousand Oaks, CA: SAGE.

Ermine, W. (1999). Aboriginal epistemology. In M. Battiste (Ed.), *First Nations education
in Canada: The circle unfolds* (pp. 101-12).Vancouver: UBC Press.

Escobar, A. (1995). *Encountering development: The making and unmaking
of the third world.* Princeton, NJ: Princeton University Press.

Estrada, Jiménez V. M. (2005). The Tree of Life as a research methodology.
The Australian Journal of Indigenous Education, 34, 44-52.

Eversole, R. (2005). Who vision? Whose rules? A culture and development typology.
Canadian Journal of Development Studies, 26(2), 295-308.

Ferguson, J. (1994). *The anti-politics machine: "Development," depoliticization,
and bureaucratic power in Lesotho*. Cambridge: Cambridge University Press.

Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). A guide to conducting
culturally-responsive evaluations. In J. Frechtling, (Ed.), *The 2010 user-friendly handbook
for project evaluation* (pp. 75-96). Arlington, VA: National Science Foundation.

Gariba, S. (2007, January). Changing paradigms in participation & partnership for
development evaluation in Africa: Are we ready for the race? 4th Conference of the African
Evaluation Association. Niamey, Niger.

Grande, S. (2008). Red pedagogy: The un-methodology. In N. K. Denzin, Y. S. Lincoln,
& L. T. Smith (Eds.), *Handbook of critical and Indigenous methodologies* (pp. 233-254).
Thousand Oaks, CA: SAGE.

Hermes, M. (1998). Research methods as a situated response: Towards a First Nations'
methodology. *Qualitative Studies in Education, 11*(1), 155-168.

Hood, S. (1998). Responsive evaluation Amistad style: Perspectives of one African-
American evaluator. In R. Sullivan (Ed.), *Proceedings of the Stake Symposium
on Education Evaluation.* Urbana-Champaign, IL: University of Illinois Press.

Hood, S. (2001). Nobody knows my name: In praise of African American evaluators
who were responsive. In J. C. Greene & T. A. Abma (Eds.) *Responsive Evaluation,
New Directions for Evaluation*, Number 92 (pp. 31-43). San Francisco: Jossey-Bass.

Hood, S., Hopson, R., & Frierson, H. (2005). *The role of culture and cultural context:
a mandate for inclusion, the discovery of truth and understanding in evaluative theory
and practice*. Greenwich, CT: Information Age Publishing.

Hopson, R. (2003). The problem of the language line: Cultural and social reproduction of
hegemonic linguistic structures for learners of African descent in the USA. *Race, Ethnicity
and Education, 6*(3), 227-245.

Kawakami, A. J., Aton, K., Cram, F., Lai, M. K., & Porima, L. (2008). Improving the practice of evaluation through indigenous values and methods: Decolonizing evaluation practice – Returning the gaze from Hawai'i and Aotearoa. In N. L. Smith & P. R. Brandon (Ed.), *Fundamental issues in evaluation* (pp. 219-242). New York, NY: Guilford.

Kirkhart, K. E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice, 16*(1), 1-12.

Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 21-39). Greenwich, CT: Information Age Publishing.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400-413.

Kovach, M. (2010). *Indigenous methodologies; Characteristics, conversations, and contexts*. Toronto, Ontario, Canada: University of Toronto Press.

LaFrance, J., & Nichols, R. (2009). Indigenous Evaluation Framework: Telling our story in our place and time. Alexandria, VA: American Indian Higher Education Consortium (AIHEC).

LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous Evaluation Framework. *The Canadian Journal of Program Evaluation, 23*(2), 12-31.

LaFrance, J., Nichols, R., & Kirkhart, K. E. (forthcoming). Culture writes the script: On the centrality of context in Indigenous evaluation. In R. F. Conner, J. Fitzpatrick, & D. J. Rog (Eds.), *Context: A framework for its influence on evaluation practice. New Directions for Evaluation.* San Francisco: Jossey-Bass.

Lee, S. A., & Farrell M. (2006). Is cultural competency a backdoor to racism? *Anthropology News, 47*(3): 9-10.

Little Bear, L. (2000). Jagged worldviews colliding. In. M. Battiste (Ed.), *Reclaiming Indigenous voice and vision* (pp. 77-85). Vancouver, BC: UBC Press.

Madison, A. M. (Ed.) (1992). *Minority issues in program evaluation, New Directions for Program Evaluation*, Number 52, San Francisco: Jossey-Bass.

Mertens, D. M. (2008). *Transformative research and evaluation*. New York, NY: Guilford Press.

Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. In D. Huffman & F. Lawrenz (Eds.), *Critical issues in STEM evaluation, New Directions for Evaluation*, Number 109 (pp. 35-51). San Francisco: Jossey-Bass.

Mertens, D. M., & Wilson, A.T. (2012). Program evaluation theory and practice: A comprehensive guide. New York, NY: The Guilford Press.

Mishler, E. G. (1979). Meaning in context: Is there any other kind? *Harvard Educational Review, 49*(1), 1-19.

Patton, M. Q. (Ed.) (1985). *Culture and evaluation, New Directions for Program Evaluation*, Number 25. San Franciso: Jossey-Bass.

Pronk (2009). Development, culture, and conflict. *Development, 52*(4), 445-455.

Rodney, W. (1981). *How Europe underdeveloped Africa*. Washington, DC: Howard University Press.

Samuels, M., & Ryan, K. (2011). Grounding evaluations in culture. *American Journal of Evaluation, 32*(2), 183-198.

Smith, L. T. (1999). Decolonizing methodologies; Research and Indigenous peoples. London: Zed Books, Ltd.

Smith, G. H. (2004). Mai i te Maramatanga, kit e Putanga Mai o te Tahuritanga; From Conscientization to Transformation. *Educational Perspectives: Journal of the College of Education/ University of Hawai'i at Manoa, 37*(1), 46-52.

Stanfield, J. H., II (2011). Epistemological reconsiderations and new considerations: Or what I have been learning since 1993. In J. H. Stanfield, II (Ed.), *Rethinking race and ethnicity in research methods* (pp. 11-26). Walnut Creek, CA: Left Coast Press.

Tafoya, T. (1995). Finding harmony: Balancing traditional values with Western science in therapy. *Canadian Journal of Native Education, 21* (supplement), 7-27.

Verhelst, T. G. (1990). *No life without roots; Culture and development*. (B. Cummings, Trans.). London: Zed Books Ltd. (Original work published 1987).

Weber-Pillwax, C. (1999). Indigenous research methodology: Exploratory discussion of an elusive subject. *Journal of Educational Thought, 33*(1), 31-45.

Williams, C. C. (2006). The epistemology of cultural competence. *Families in society, 87*(2), 209-220.

Wilson, S. (2008). *Research is ceremony: Indigenous research methods*. Halifax & Winnipeg, Canada: Fernwood Publishing.

Part 2
Methodological implications for Equity-focused evaluations[1]

---

1   For additional details on methodological implications, please see Bamberger and Segone, 2011, *How to design and manage Equity-focused evaluations* , UNICEF

METHODOLOGICAL ISSUES FOR DESIGNING AND IMPLEMENTING EQUITY-FOCUSED EVALUATIONS[1]

Michael Bamberger, Independent consultant

Introduction

Segone's article in this book defined equity and why it is important for the planning and implementation of development programs. It also discussed why Equity-focused evaluations are needed. The present article discusses strategies and methods for evaluating how well development interventions address and achieve equity objectives. There are a number of references to the Equity-focused evaluations resource centre available at www.mymande.org where methodological issues are discussed in more detail.

Evaluating the impact of Equity-focused interventions

It is useful to distinguish between *simple* equity-focused projects, and *complex* equity-focused policies and other national level interventions as these affect the available Equity-focused evaluation design options.

"Simple" equity-focused projects. The term "simple" refers to the scope, relative clarity in the definition of objectives and the organization of the project. They tend to follow a standard, blueprint approach, with a small number of components, usually with a defined budget and target population and clearly defined objectives. They are often based on a logic or program theory model laying out the objectives, stages of implementation and performance indicators. A simple equity-focused project is a relatively "closed" system in which funding agencies have significant control over the inputs and results. "Simple" does not mean that the project is "easy" to implement or that there is a high probability of success. Many "simple" projects operate in poor and vulnerable communities with high levels of insecurity and conflict and there are many challenges to achieving project objectives.

---

1   Based upon Bamberger M. and Segone M. (2011). *How to design and manage Equity-focused Evaluation*, UNICEF.

Complex equity-focused policies are more difficult to characterize as there are many different scenarios, but they will often have the following features: a number of different components and often distinct programs, usually large-scale, involving a number of different stakeholders, often with no clear definition of the range of services or the target population, and often they do not have a defined start and end date. The evaluation of complex equity-focused policies requires the use of more creative and less quantitatively oriented evaluation methodologies than those used in "simple" project Equity-focused evaluations.

Causality, attribution and the importance of the counterfactual

The assessment of impacts or causality requires an estimate of what would have been the condition of the target population if the intervention had not taken place. In order to control for the influence of other factors that might contribute to the observed changes, it is necessary to define a *counterfactual*. For experimental and quasi-experimental designs, the counterfactual is estimated through a comparison group that is statistically or judgmentally matched to the target population. In the real-world, it has only proved possible to use statistical comparison groups in a small proportion of interventions, so one of the main challenges for Equity-focused evaluations is how to define a credible alternative counterfactual to answer the question "what would have been the situation of the worst-off groups if the intervention had not taken place"?

Evaluating equity-focused impact at the policy level

Systems approach to evaluation

Systems approaches offer potentially valuable ways to understand how a particular intervention is affected by, and in turn can influence the public and private service delivery systems within which program implementation takes place (see Williams 2005, and Williams and Imam 2007). Systems approaches can be particularly helpful for evaluating equity focused programs as many of these operate within, and seek to change systems which are often resistant to accepting (or are actively opposed to) the proposed focus on the worst-off sectors of society. Some of the systems approaches that can potentially be applied to Equity-focused evaluations include: the Systems Dynamics Approach, Soft Systems Methodology (SSM),

and Cultural Historical Activity Theory (CHAT). See Williams and Imam (2007) for an overview of these approaches.

Systems analysis introduces some radically different ways of thinking about evaluation, all of which are potentially important for Equity-focused evaluations. Some of the insights that can be drawn from these approaches include:

- Development programs and policies are embedded into an existing social system with its own historical traditions, linkages among different beneficiaries, actors and owners. The intervention must adapt to, and will often be changed by the existing systems.

- Different actors may have very different perspectives on how the new intervention operates and even whether it is accepted at all, will be affected by these perspectives.

- Systems have boundaries (which may be open or closed) which affect how widely the new intervention will be felt.

- New interventions create contradictions and often conflicts and the program outcomes will be determined by how these conflicts are resolved.

Unpacking complex policies and other national level interventions into components that can more easily be evaluated

National-level policies usually have several components with different objectives and each organized differently. Furthermore, as the policies are intended to operate throughout the whole country, it is difficult to define a conventional comparison group. However, it is often possible to "unpack" the policy into distinct components each of which can be evaluated separately:

- It might be possible to break down an education reform program designed to increase girls' access to education into: financial incentives to schools to enroll and graduate more girls; scholarships; transport vouchers and gender-sensitive textbooks. Many of these components could be evaluated using conventional evaluation designs.

- Many national level policies will have measurable outcomes at provincial and local levels.

- Many national policies are implemented in phases, or do not reach all areas at the same time. Consequently it is often possible to use Pipeline designs (see below) to identify comparison areas that have not yet been affected by the intervention.

Pipeline evaluation designs

Pipeline designs take advantage of the fact that some policy and national-level interventions are implemented in phases (either intentionally or due to unanticipated problems). Consequently the areas, districts or provinces where the intervention has not yet started can be used as a comparison group. However, it is important to determine why certain regions have not yet been included and to assess how similar they are to regions already covered. When there is a systematic plan to incorporate different provinces or districts in phases, the design may work well, but when certain regions have been unintentionally excluded due to administrative or political problems the use of the pipeline design may be more problematic.

Policy gap analysis

Policy gap analysis identifies key policy priorities and target groups and assesses how adequately policies address these priorities. It reviews the whole spectrum of public policies to identify limitations of individual policies as well as problems arising from a lack of coordination between different policies. This analysis is particularly important for equity issues because inequities have multiple causes and require a coordinated public sector approach, and often the worst-off groups fall through gaps in the social safety net.

The analysis is normally conducted at the national level using secondary data from surveys and agency records and complemented by desk reviews, consultation with key informants, focus groups and possibly visits to ministries or service delivery centers. Techniques such as quintile analysis are used to identify the worst-off groups and to compare them with other groups on indicators such as school enrolment or use of health services.

Bottleneck Analysis, Knowledge, Attitude and Practice (KAP) studies and Citizen Report Cards (see below) can provide additional information. It is sometimes possible to incorporate a special module into survey conducted by another agency to fill in some of the information gaps.

Using other countries as the comparison group

When evaluating national level policies it is sometimes possible to use other countries as a comparator. Extensive comparative international data is now available from sources such as Living Standards Measurement Surveys (LSMS), Demographic and Health Surveys (DHS) and Multiple Indicators Cluster Survey (MICS). Over the past few years data bases are also becoming available on governance and topics such as political and community participation and corruption. These data bases permit comparison with a large sample of countries with similar socio-economic and other relevant characteristics. It is however more difficult to find comparative data specifically relating to worst-off groups.

Concept mapping

Concept mapping interviews stakeholders or experts to estimate policy effectiveness, outcomes or impacts (Kane and Trochim 2007, see also the resource center available at www.mymande.org). It is well suited as a tool for Equity-focused evaluations as it allows experts to use their experience and judgment to help define the dimensions that should be used to evaluate equity policies and then to rate policies on these dimensions. This is particularly useful for evaluating equity-focused policies where objective quantitative indicators are difficult to apply. A comparison of the average ratings for areas receiving different levels of intervention combined with a comparison of ratings before and after the intervention can provide a counterfactual. Concept mapping could be used to evaluate programs promoting, for example, gender mainstreaming, or policies that seek to increase access of worst-off groups to public services, to provide them with equal treatment under the law, or that protect them from violence and other sources of insecurity.

Portfolio analysis

Many complex equity-focused policies include many different interventions that contribute to the policy objective in different ways. Portfolio analysis classifies relevant interventions into performance areas, identifies and assesses the kinds of information available on these projects such as: the existence of a logic model, monitoring data, ratings of quality at entry, implementation and completion, and other kinds of evaluation reports. See the resource center for an example of how portfolio analysis was used to evaluate effectiveness of World Bank support for gender and development.

Projects are rated on different dimensions and summary indicators are produced for all of the projects in each performance area.

Average ratings are computed for each dimension, and combined to obtain an overall assessment for each performance area. Many agencies use the OECD/DAC evaluation criteria (see resource centre), for these overall assessments. Additional criteria relevant to humanitarian settings, such as coherence, integration and coverage, may also be used. If resources permit, a sample of projects may be selected for field studies to compare the data from these secondary sources with experience on the ground.

Evaluating equity-focused impact at the project and program levels

Conventional project level statistical evaluation designs estimate the contribution of an intervention (project) to the observed changes in an outcome indicator by identifying a comparison group with similar characteristics to the project population but that does not have access to the intervention. If there is a statistically significant difference in the degree of change between the two groups, this is taken as evidence of a potential project effect. The three main methods for matching, in descending order of statistical precision are: randomized control trials, quasi-experimental designs with statistical matching of the project and comparison groups and quasi-experimental designs with judgmental matching. The strength of the statistical analysis is influenced by how closely the project and comparison groups are matched, the sample size and the size of the change in outcomes (effect size). A careful evaluator will use triangulation (obtaining independent estimates on the causes of the changes from secondary data, key informants, direct observation or other sources) to check the estimates.

The resource centre presents a list of 7 basic impact evaluation designs and also an expanded list with 20 evaluation design options. These take into consideration how the comparison group was selected and how the baseline conditions of the project and comparison groups were estimated. Many evaluations do not begin until late in the project cycle so it is important to be familiar with techniques for reconstructing baseline data (Bamberger, 2010).

Estimating project impacts using non-experimental designs (NEDs)

Non-experimental designs (NEDs) do not include a statistical counterfactual so it is not possible to control statistically for factors that might have produced the changes in the output indicators. While

NEDs are often considered the default option when budget con-
straints do not permit the use of a comparison group, there are
many situations where a NED is the methodologically strongest
evaluation design, for example when:

- the program involves complex processes of behavioral change;
- outcomes are not known in advance;
- outcomes are qualitative and difficult to measure;
- different local settings that are likely to affect outcomes;
- it is important to understand the implementation process;
- the project evolves slowly over a relatively long period of time;

Examples of potentially strong NEDs include (see resource center
for examples of each approach):

- *Single case analysis.* Changes of behavior or performance are
  compared for a single case (individual or group) before and
  after the intervention. The baseline is usually an assessment
  by a group of experts of pre-treatment behavior. The treatment
  is applied at least three times, and if a significant change is
  observed on each occasion then the treatment is considered to
  have been effective. The experiment would then be conducted
  again in different settings to build up data on when and why the
  treatment works.

- *Longitudinal designs.* The subject group is observed over a
  long period of time to describe the process of change and how
  this is affected by contextual factors in the local setting. These
  approaches are useful for understanding, for example changing
  relations between spouses as a result of a program intervention;
  or to monitor changes in gender-based community violence.

- *Interrupted time series.* The design can be used when periodic
  data is available over a long period of time, beginning before the
  intervention takes place and continuing after the intervention.
  The analysis examines whether there is a break in the intercept
  or the slope at the point where the intervention took place.

- *Case study designs.* A sample of case studies is selected to
  represent the different categories or typologies of interest to the
  evaluation. The cases describe how different groups respond to
  the project intervention and this provides an estimate of project
  impacts.

Analytical models for Equity-focused evaluations

Theory-based Equity-focused evaluations

A well-articulated theory model is important for Equity-focused evaluations at both policy and program levels. Equity interventions achieve their objectives through the promotion of behavioral change and the process of implementation, and the context have a significant impact on the accessibility of the services to difficult-to-reach groups. It is also important to understand how effectively service delivery systems have adapted to the special challenges of reaching worst-off groups. A well-articulated program theory model can:

- Define the nature and causes of the problem the program is intended to address and proposed solutions;

- Incorporate lessons from similar programs;

- Identify the intended outcomes and impacts and describe how outcomes are to be achieved;

- Define the key assumptions and hypotheses on which the program design is based;

- Identify the contextual factors likely to affect implementation and outcomes;

- Identify the main risks and reasons why the program may not achieve its objectives.

The model also helps interpret the evaluation findings. If intended outcomes are not achieved, the model can trace-back through the steps of the results chain to identify where actual implementation experience deviated from the original plan. If implementation experience and outcomes correspond reasonably closely to the program design, this provides prima facie evidence to attribute the changes to the results of the program. However, there may be other plausible explanations of the changes, so a well-designed program theory should define and test rival hypotheses.

The bottleneck analysis framework

Bottleneck supply and demand analysis provides a framework for the analysis of factors affecting the access of worst-off groups to public services, and for identifying the strengths and weaknesses of different service delivery systems (the framework is presented

in more detail in pp. 45-50 of "How to design and manage equity-focused evaluations" at www.mymande.org).

The framework assesses four components:

• Use of services by worst-off groups.

• Supply side factors. The adequacy of budgets and other resources, the efficiency of service delivery, the cultural responsiveness of targeting and implementation, and the participation of worst-off groups in program planning and implementation.

• Demand-side factors. The behavioral changes required for the program to operate successfully by the target worst-off groups, the service delivery agency and policymakers and planners. Have logistical and cultural factors constraining effective demand (distance, time, cost, availability of transport, cultural constraints) been addressed?

• Understanding and addressing contextual factors (political, economic, organizational, socio-cultural, natural environmental factors) that affect program success.

Tools for Equity-focused evaluations

When designing an Equity-focused evaluation it is important to understand the context and the factors affecting implementation and accessibility to different worst-off groups. It is also important to understand perceptions and attitudes of implementing agencies and society towards different worst-off groups. In most situations it will be useful to conduct a rapid diagnostic study to understand the intervention and its context. For a small intervention in only a few locations, it may be possible to conduct the diagnostic study in a few weeks; for a large and widely dispersed intervention significantly more time may be required. The following are some of the kinds of information that the diagnostic study will collect:

• How are the problems the intervention is designed to address, currently addressed?

• What are the opinions of different sectors of the community concerning these services?

• Have similar projects been tried earlier and how did they work out?

- Which groups are most affected by the problems to be addressed?

- What are the reasons for lack of access of different groups?

Diagnostic studies will normally use one or more of the following data collection methods: participant and non-participant observation, rapid household surveys, interviews with key informants and local experts and focus groups.

Knowledge, Attitude and Practice (KAP) studies

KAP studies assess the effectiveness of information campaigns and service delivery program by examining: *Knowledge* about the intervention; *attitudes* towards the proposed services and how they will be delivered; and *behavior* with respect to acquiring and using the services or adopting the recommended practices. KAP studies have been widely used to assess family planning and public health programs, agricultural extension, and HIV/AIDS prevention programs (among others).

A KAP study is usually implemented in 5 steps: Domain identification (Defining the intervention, the knowledge to be communicated, the attitudes to be measured and the indicators of acceptance and use); Identifying the target audience; Defining the sampling methods; Defining the data collection procedures; and Analysis and reporting.

Citizen report cards

Citizen report cards use surveys covering a major urban area asking households which public service agencies (education, health, police, transport, water etc.) they have contacted within the last 12 months to address a particular problem. They are asked: were they able to resolve their problem; how many visits were required; how were they treated by agency staff; and did they have to pay bribes. Average ratings are calculated for each agency on each dimension. The surveys may be repeated (usually 2-3 years later) to measure changes in performance. Samples can be designed to over-sample worst-off populations (for example, the citizen report cards in Bangalore included a separate stratum for slum dwellers).

The credibility and independence of the research agency is critical as a common reaction of public service agencies is to deny the representativity of negative findings and to challenge the professional competence or motives of the research agency. It is important to have a sufficiently large sample to be able to disaggregate the data to study different worst-off groups.

Cost-effectiveness analysis

Cost-effectiveness analysis compares different services or delivery systems in terms of their costs and results (Levin and McEwan 2001). The analysis may compare average costs of service delivery and costs for reaching special groups (e.g. worst-off groups), or it may compare alternative delivery systems for reaching special groups. Some of the key elements in cost-effectiveness analysis include:

- Ensuring that the services to be compared are equivalent.

- Identifying all of the costs of the programs being compared and ensuring that they are measured in an equivalent way and that any hidden subsidies are identified and monetized.

- A final issue concerns the question of scaling-up. If a pilot is considered successful, replication on a larger scale will often be considered. However, it is difficult to estimate how scale will affect costs. So care must be taken when assuming that because a small program is relatively inexpensive the same will be true on a larger scale.

Public expenditure tracking (PETS) studies

PETS studies track the percentage of budget funds approved for services such as schools and health clinics that actually reach these agencies (Bamberger and Ooi 2005). The studies involve a careful review of disbursement procedures, combined with interviews with agency staff to track the flow of funds, to note the delay in transfer from one level to another and the proportion of funds that get lost at each stage. If data is available it would be possible to track the proportion of funds that reach worst-off groups.

Public expenditure Benefit Incidence Analysis (BIA)

BIA estimates the effectiveness with which public expenditures in sectors such as health and education reach worst-off groups. Normally the analysis estimates access to services by income quintile and it is rarely possible to examine other dimensions of inequity (such as female-headed households, and families with disabled children). The analysis requires three types of data: government spending on a service; public utilization of the service; and the socioeconomic characteristics of the population using the service (Davoodi, Tiongson and Asawanuchit, 2003).

The analysis can be repeated to assess the effects of new legislation or external factors such as a financial crisis, on expenditure

incidence. BIA has been used extensively in the preparation of national poverty reduction strategies, but it is a potentially useful tool for Equity-focused evaluations. Two limitations of BIA for Equity-focused evaluations are that data is not available on the quality of services, and data is only available at the level of the household so it is not possible to examine access to services by vulnerable groups such as children, or the physically and mentally handicapped.

Mixed methods data collection and analysis

Evaluation strategies such as bottleneck analysis and KAP studies require a mixed method design that combines quantitative (QUANT) methods that permit unbiased generalizations to the total population, analysis of the distribution of sample characteristics and breakdown into sub-groups, and testing for statistically significant differences between groups; with qualitative (QUAL) methods describing the lived-experiences of individual subjects, groups or communities, examining complex relationships and explaining how program effectiveness is affected by the context in which the program operates.

One of the key strengths of mixed methods is that sample designs permit the selection of cases for in-depth analysis that are statistically representative of the population from which the cases are selected. Credible methods of sample selection are critical because case studies often uncover issues or weaknesses in service delivery (such as sexual harassment, lack of sensitivity to different ethnic groups, or corruption). If it is not possible to show that the findings are representative, it is easy for agencies to dismiss negative findings as not being typical. Mixed method is an integrated evaluation approach with a unique approach to each stage of the evaluation.

Hypothesis development. Mixed method designs combine QUANT *deductive* hypotheses (pre-existing theories are used to define hypotheses that will be tested) and QUAL *inductive* hypotheses that emerge during data collection and are often revised as the evaluation progresses. For example, there may be a *deductive* hypothesis about the effects of economic downturns on the magnitude and distribution of inequity which can be tested statistically with survey data. At the same time field work might identify differences in how ethnic or age groups respond to economic downturns and corresponding hypotheses can be formulated *inductively*.

Sampling. QUANT random sampling can be combined with purposive QUAL sampling to generalize from the in-depth case studies or interviews to the total sample population.

Data collection. Mixed methods combine quantitative methods such as surveys, aptitude tests, and anthropometric measures; with QUAL methods such as observation, in-depth interviews, and the analysis of artifacts.

Data analysis. Mixed methods can combine QUANT and QUAL data analysis methods in the following ways:

- Parallel analysis: QUAL and QUANT data are analyzed separately using conventional analysis methods.

- Conversion of QUAL data into a numerical format or vice versa.

- Sequential analysis: QUANT analysis followed by QUAL analysis or vice versa.

- Multi-level analysis.

Mixed method designs can be considered as a continuum ranging from mainly QUANT designs with a small QUAL component; through balanced designs, to mainly QUAL designs with a small QUANT component. There are three main kinds of mixed methods design:

- *Sequential*: The evaluation either begins with QUANT data collection and analysis followed by a QUAL data collection and analysis or vice versa. Designs can also be classified according to whether the QUANT or QUAL components of the overall design is dominant.

- *Parallel*: The QUANT and QUAL components are conducted at the same time.

- *Multi-level*: The evaluation is conducted on various levels at the same time. Multi-level designs are particularly useful for studying the delivery of public services such as education, health, agricultural extension where it is necessary to study both how the program operates at each level and the interactions between levels.

Triangulation, which is an integral part of mixed method designs, combines two or more independent sources to assess the validity of data that has been collected and to obtain different interpretations of what actually happened during project implementation and

the effects on different sectors of the population. Triangulation can compare information: collected by different interviewers, collected at different times or in different locations, and obtained using different data collection methods.

Reconstructing baseline data when the evaluation is not commissioned until late in the implementation cycle

There are many situations where evaluators do not have access to baseline data, either because it was not collected or because it was of poor quality. There are a number of strategies that can be used to "reconstruct" baseline data for Equity-focused evaluations (Bamberger, 2010):

- Using data from the program M&E system and administrative records.

- Using records from other similar programs (schools, health clinics etc.) to construct a comparison group.

- Using records from national data sets (MDGs, LSMS, DHS, MICS etc.).

- Using recall: respondents are asked to recall their situation or that of their community at the time the project began. For example, respondents can be asked to recall their income or expenditures, time spent travelling to work or collecting water, which children attended school before the project school was built. While recall is often the only available source of information on the past it is usually difficult to detect potential sources of bias (problems with memory, difficulties of locating events in time or intentional distortion). In most cases there are also no guidelines to estimate and adjust for the direction and magnitude of recall bias.

- Key informants.

- Focus groups and PRA.

All of these techniques can be used for reconstructing baseline estimates for the number and types of worst-off populations and the particular problems they faced. However, it is more difficult to obtain reliable estimates on the subtle and sensitive concepts relating to equity than to obtain information on things like school enrolment and travel time.

Potential challenges in promoting and implementing Equity-focused evaluations

The following are some of the challenges that organizations and countries may face in promoting and implementing Equity-focused evaluations:

- Government reluctance to accept the disaggregated indicators produced by Equity-focused evaluations which can show country performance in a poorer light.

- Political and social resistance to addressing the causes of exclusion and vulnerability and to empowering worst-off groups.

- Lack of interest and incentives to invest resources in worst-off groups.

- Governance: many government agencies do not have the capacity to design and implement targeted programs for worst-off groups. In countries where corruption is an issue, worst-off groups, are those with the least ability to defend their rights.

- Methodological challenges in the evaluation of complex interventions. Equity-focused interventions are often more complex, and consequently must use more advanced methodologies.

- Lack of disaggregated data or data collection capacity, and reluctance to change existing methodologies.

- Additional cost and complexity.

- Reluctance of some governments to work with civil society. In many countries civil society organizations have the greatest experience using the qualitative and mixed method designs required for Equity-focused evaluations. However, governments are sometimes reluctant to work with these organizations as they are perceived as being critical of government or wishing to address sensitive issues such as gender equality or the situation of refugees.

References

Bamberger M. and Segone M. (2011). *How to design and manage Equity-focused evaluations.* UNICEF. Available at: http://www.mymande.org/?q=virtual

Bamberger, M (2010). *Reconstructing baseline data for impact evaluation and results measurement. No. 4.* The Nuts and Bolts of M&E Series. Poverty Reduction and Equity Department. The World Bank. Available at http://siteresources.worldbank.org/INTPOVERTY/Resources/335642-1276521901256/premnoteME4.pdf.

Bamberger, M; Rao, V and Woolcock, M (2010) Using mixed methods in monitoring and evaluation: Experiences from international development. In Tashakkori, A and Teddlie, C. Sage Handbook of Mixed Methods in Social and Behavioral Research. Sage Publications. pp 613-42

Bamberger, M; J. Rugh and L. Mabry (2006). Real World Evaluation. Sage Publications.

Bamberger and Ooi (editors) 2005. *Influential Evaluations: Detailed case studies*. Independent Evaluation Group. World Bank. Available at www.worldbank.org/oed/ecd

Davoodi, Tiongson and Asawanuchit (2003). *How useful are benefit incidence analyses of public education and health spending?* IMF Working Paper. International Monetary Fund.

Levin, Henry and Patrick McEwan (2001). *Cost-effectiveness Analysis: Methods and Applications*. Sage Publications.

Rogers, E (2003). *Diffusion of Innovations*. Free Press.

Williams, B (2005) *Systems and Systems Thinking* in Mathison, S (editor) Sage Encyclopedia of Evaluation pp. 405-412.

Williams, B and Imam, I (eds), (2007). *Systems Concepts in Evaluation: An expert anthology*. American Evaluation Association.

DEVELOPMENTAL EVALUATION FOR EQUITY-FOCUSED EVALUATIONS

*Michael Quinn Patton,
Founder and Director, Utilization-Focused Evaluation*

Developmental evaluation supports innovative intervention *development* to guide adaptation to emergent and dynamic realities in complex environments (Patton, 2011b). Developmental evaluation is *utilization-focused* (Patton, 2011b) in that it focuses on a specific intended use – development – for specific intended users: social innovators adapting their interventions in complex dynamic environments. Evaluation for equity and the fostering of human rights, as part of achieving meaningful development results, often occurs in complex adaptive systems. A complex system is characterized by a large number of interacting and interdependent elements in which there is no central control. Complex environments for social interventions and innovations are those in which what needs to be done to solve problems is uncertain, where key stakeholders are in conflict about how to proceed. This is typically the situation when fostering human rights. What has worked in one place may not work in another. Context matters. Variations in culture, politics, resources, capacity, and history will affect how development initiatives unfold and how attention to equity and human rights is incorporated into those initiatives. In such situations, informed by systems thinking and a sensitivity to complex nonlinear dynamics, developmental evaluation supports increased effectiveness of interventions, social innovation, adaptive management, and ongoing learning.

The developmental evaluator is often part of a development team whose members collaborate to conceptualize, design, and test new approaches in a long-term, on-going process of continuous development, adaptation, and experimentation, keenly sensitive to unintended results and side effects. The evaluator's primary function in the team is to infuse team discussions with evaluative questions, thinking, and data, and to facilitate systematic data-based reflection and real-time decision-making in the developmental process.

Improvements versus developments

There are many approaches to evaluation. Each, including developmental evaluation (DE), fulfills a specific purpose and adds a particular

kind of value. As noted above, DE has proven especially relevant and attractive to those interested in systems change and social innovation in complex dynamic systems. These are systems where people are trying to bring about major social change by fighting poverty; homelessness; inequity; human rights abuses; community and family violence; and helping people with AIDS, severe disabilities, chronic diseases, and victims of natural disasters and war. A deep commitment to fostering human rights and supporting equity undergirds many of these interventions and systems-change initiatives. Canadian colleagues Frances Westley, Brenda Zimmerman, and I have studied successful social innovations. We reported what we found in a book entitled *Getting to Maybe: How the World Is Changed* (Wesley, Zimmerman, & Patton, 2006). To be a change agent is to think boldly and envision grandly. Complexity theory shows that great changes can emerge from small actions. This involves a belief in the possible, even the "impossible". Moreover, major social innovations don't follow a simple linear pathway of change. There are ups and downs, roller coaster rides along cascades of dynamic interactions, unexpected and unanticipated divergences, tipping points and critical mass momentum shifts, and things often get worse before they get better as systems change creates resistance to and pushback against the new.

Traditional evaluation approaches are not well-suited for such turbulence. Traditional evaluation aims to control and predict, bring order to chaos, by carefully specifying and measuring fidelity of implementation and attainment of predetermined priority outcomes. In contrast, developmental evaluation accepts turbulence and uncertainty as the way of the world, as social innovation unfolds in the face of complexity. Developmental evaluation adapts to the realities of complex non-linear dynamics rather than trying to impose order and certainty on a disorderly and uncertain world. DE does this by tracking and documenting emergent and dynamic implementation adaptations and results.

Many of those working to foster human rights tell me that they have experienced evaluation methods that are entirely unrelated to the nature of their initiatives. Identifying clear, specific, and measurable outcomes at the very start of an innovative project, for example, may not only be difficult but also counter-productive. Under conditions of great uncertainty, outcomes can emerge through engagement, as part of the process for change rather than prior to such change efforts. So-called "SMART objectives,"[1] imposed prema-

---

1   SMART objectives: specific, measureable, attainable, realistic, timely

turely, are not smart – and can, in fact, do harm by limiting responsiveness and adaptability. Developmental evaluation is designed to be congruent with, and to nurture developmental, emergent, innovative, and transformative processes.

Developmental evaluation and Complexity theory

Complexity as a construct is a broad tapestry that weaves together several threads relevant to innovation and evaluation: non-linearity; emergence; dynamic systems; adaptiveness; uncertainty; and co-evolutionary processes (Patton, 2011a). Developmental evaluation, likewise, centers on situational sensitivity, responsiveness, and adaptation, and is an approach to evaluation especially appropriate for situations of high uncertainty, where what may and does emerge is relatively unpredictable and uncontrollable. Developmental evaluation tracks and attempts to make sense of what emerges under conditions of complexity, documenting and interpreting the dynamics, interactions, and interdependencies that occur as innovations and systems-change processes unfold.

Complex adaptive systems

Complexity writings are filled with metaphors that try to make complex phenomena understandable to the human brain's hard-wired need for order, meaning, patterns, sense-making, and control, ever feeding our illusion that we know what's going on. We often don't. But the pretense that we do is comforting – and sometimes necessary for some effort at action. So complexity theorists talk of flapping butterfly wings that change weather systems and spawn hurricanes; individual slime molds that remarkably self-organize into organic wholes; ant colonies whose frantic service to the Queen mesmerize us with their collective intelligence; avalanches that reconfigure mountain ecologies; bacteria that *know* the systems of which they are a part without any capacity for self-knowledge; and 'black swans' that appear suddenly and unpredictably to change the world. Complexity science offers insights into the billions of interactions in the global stock market; the spread of disease throughout the world; volatile weather systems; the evolution of species; large scale ecological changes; and the flocking of migrating birds. Complexity theorists explain the rise and fall of civilizations, and the rise and fall of romantic infatuation. That's a lot of territory. It can and should include attention to the rise and fall of evaluations.

Dealing with the unexpected

There is a lot of lip service in evaluation about looking for unanticipated consequences and assessing side effects; in reality, these are typically token elements of evaluation designs and inadequately budgeted, which are rarely given serious time and attention because of the overwhelming focus on measuring attainment of intended outcomes and tracking the preconceived performance indicators. You have to go out into the real world, do fieldwork, engage in open inquiry, talk to participants in programmes, and observe what is going on as interventions and innovations unfold to detect unanticipated consequences. I find that evaluators typically approach the unexpected and unanticipated in a casual and low-priority way, essentially saying, we'll look for unanticipated consequences and emergent outcomes if we have time and resources after everything else is done. But, of course, there seldom is time or resources. But the probabilities for unexpected impacts become quite high under conditions of complexity and so, developmental evaluators make expecting the unexpected fundamental to the work at hand.

Developmental Evaluation and learning

Developmental evaluation supports learning to inform action that makes a difference. This often means changing systems, which involves getting beyond surface learning to a deeper understandings of what is happening in a system. Social innovators and social entrepreneurs, especially those working on issues of human rights and equity, are typically trying to bring about fundamental changes in systems, to change the world. To do so, they have to understand how the system they want to change is operating and to make the changes that change the system itself, by getting beyond temporary and surface solutions. This involves *double-loop learning*.

For decades three stories have been endlessly repeated: one about the stream of ambulances at the bottom of the cliff instead of building fences at the top; one about the numerous dead bodies coming down the river but all we do is build more impressive services for fishing them out; and one about giving someone a fish versus the value of teaching that person how to fish. In reviewing these stories, distinguished Australian action research scholar and practitioner Yolande Wadsworth (2011), has commented that they are reminders about our repeated tendency to go for the short-term quick fix (rather than to examine, come to understand, and take action to change how a system is functioning), that creates the very

problems being addressed. Double-loop learning involves systemic solutions and is supported by evaluation attuned to looking for system explanations, and offering systemic insights.

In single-loop learning, people modify their actions as they evaluate the difference between desired and actual outcomes, and make changes to increase attainment of desired outcomes. In essence, a problem-detection-and-correction process, like formative evaluation, is *single-loop* learning. In double-loop learning, those involved go beyond the single loop of identifying the problem and finding a solution to a second loop that involves questioning the assumptions, policies, practices, values, and system dynamics that led to the problem in the first place, and then intervening in ways that involve the modification of underlying system relationships and functioning. Making changes to improve immediate outcomes is single loop learning; making changes to the system to prevent the problem or embed the solution in a changed system, involves double-loop learning. *Triple-loop learning* involves learning how to learn, and is embedded in the processes of developmental evaluation.

Developmental Evaluation in the context of Development Evaluation

Developmental evaluation is easily confused with development evaluation. They are not the same, though developmental evaluation can be used in development evaluations.

Development evaluation is a generic term for evaluations conducted in developing countries, usually focused on the effectiveness of international aid programmes and agencies. *The Road to Results: Designing and Conducting Development Evaluations* (Imas and Rist, 2009) is an exemplar of this genre, a book based on The World Bank's highly successful International Programme for Development Evaluation Training (IPDET) which the book's authors founded and direct, and on which their book is based.

Developmental evaluation, as defined and described in the *Encyclopedia of Evaluation* (Mathison, 2005, p.116), has the purpose of helping develop an innovation, intervention, programme, or systems change. The evaluator uses evaluative methods to facilitate ongoing programme, project, product, staff and/or organizational *development*. The evaluator's primary function in the team is to facilitate and elucidate team discussions by infusing evaluative questions, data and logic, and to support data-based decision-making in the developmental process.

An evaluation focused on development assistance in developing countries could use a developmental evaluation approach, especially if such developmental assistance is viewed as occurring under conditions of complexity with a focus on adaptation to local context. Developmental evaluation can be used wherever social innovators are engaged in bringing about systems change under conditions of complexity.

The 'al' in *developmental* is easily missed, but it is critical in distinguishing development evaluation from developmental evaluation.

Figure 1: DD^2 = Developmental evaluation
for development evaluation

When I first labeled and wrote about *developmental evaluation* 15 years ago (Patton, 1994), development evaluation was not a distinct and visible category of evaluation practice and scholarship. Evaluations in developing countries were certainly being conducted, but an identifiable body of literature focused on evaluating development assistance had not attracted general professional attention. One of the most important trends of the last decade has been the rapid diffusion of evaluation throughout the world including, especially, the developing world, as highlighted by formation of the International Development Evaluation Association. Confusion about the distinct and sometimes overlapping niches of development evaluation and developmental evaluation is now, I am afraid, part of the complex landscape of international evaluation. I hope that this chapter helps to sort out both the distinctions and the areas of overlap.

Examples of developmental evaluation in development contexts

• Working with agricultural scientists to take an integrated systems approach to 'orphan crops' would involve working with agronomists; soil scientists; plant breeders; water specialists; extension personnel; health; nutrition; gender researchers; and farmers, to conceptualize agricultural innovation as a complex adaptive system and identify real time indicators of the systems interactions and dynamics as the new farming approaches start to affect use of agricultural inputs, production techniques, farm

labour, and farm family dynamics. This kind of holistic intervention involves changes in how traditionally distinct agricultural and nutritional scientists engage with farmers (separately rather than together), and would affect farm family decision-making and interactions.

- A microfinance intervention examined through a developmental evaluation lens would look at the infusion of capital as triggering a leverage point in a complex adaptive system. It would have implications for a variety of business calculation and decisions; interdependencies among loan recipients; relationships with consumers; and family finances and interpersonal dynamics. Watching for and adapting to emergent outcomes beyond simple use of small loan funds would be built into the evaluation design and real time feedback, as the microcredit system developed.

Examples of developmental evaluation in Equity-focused evaluations

Developmental evaluations focusing on the marginalized and excluded populations help to adapt to rapidly changing conditions. Here are some examples.

- People living in poverty exist on the edge of subsistence. Sudden changes in food availability can move an entire population from subsistence to famine. Food insecurity can result from weather (severe drought or flood), political unrest (food transport is disrupted), and economic changes (increases in food prices). Sometimes all three factors – weather, political, and economic disruptions – occur simultaneously, creating a mutually reinforcing downward spiral on increasing desperation. Such situations require real time data about what is happening to the people affected and how well-intentioned interventions are actually performing.

- Marginalized and excluded populations are especially susceptible to contagious diseases. For example, polio immunization campaigns have to be adapted to specific development contexts. Where polio eradication efforts have floundered, as in parts of Nigeria, Pakistan, and India, new outbreaks can break-out and spread rapidly in areas where the disease was thought to have been eradicated. For example, a developmental evaluator would help monitor rumors about resistance to a vaccination campaign. Detecting and correcting such rumors in real time, as they emerge, can save lives.

- A human rights campaign anywhere in the world may have to be significantly adapted as street demonstrations calling for democratic reforms in Tunisia and Egypt (2011) change the global context within which human rights initiatives are undertaken. Marginalized, disempowered, and excluded populations can become homeless refugees when political turmoil accelerates and spreads.

- Responding to a humanitarian disaster, such as the earthquake in Haiti (2010), requires real time data about how local pockets of people are being affected; which roads are passable; where heavy rains after the earthquake are threatening the stability of remaining buildings; where there are outbreaks of cholera; where food, clean water, and medications are most desperately needed; and so on and so forth. Efforts to coordinate an international humanitarian response are inherently *developmental* because the disaster context is complex and emergent. The evaluation should also be developmental in support of ongoing humanitarian relief decision-making. Marginalized, disempowered, and excluded populations are often especially vulnerable in disaster situations because they tend to live in highly vulnerable areas that lack basic infrastructure. This makes delivering timely assistance all the more challenging. Developmental evaluation can track both developing vulnerabilities and developing interventions.

Dynamic versus static impact evaluation designs

As these examples illustrate, developmental evaluation views development interventions as dynamic and emergent in complex adaptive systems. Both the intervention and the evaluation are dynamic and adaptive. This stands in stark contrast to impact evaluations that use randomized controlled trials (RCTs) as a methodological framework. RCTs conceptualize interventions as occurring in closed systems, and study the intervention as a static and mechanical cause aimed at preconceived effects in a simple linear model of cause-effect. Such designs aim to standardize interventions and to control variation, which limits the utility and generalizability of findings. (For more on the mechanical and linear assumptions of RCTs, see Patton, 2008, chapter 12). In contrast, developmental evaluations assume that development more often occurs in complex dynamic systems and puts a premium on understanding context, real time adaptability, and ongoing development, rather than

generating high-fidelity and highly prescriptive practices. These differences go beyond methodological preferences and debates. They involve fundamentally different views about the nature of development, the contexts within which development occurs, how change occurs, and epistemological differences about what constitutes actionable knowledge.

Developmental Evaluation and accountability

The traditional approach to accountability is to evaluate whether resouces are used as planned, and whether targeted outcomes are attained. This is a static and mechanical approach to accountability that assumes designers know, three or five years in advance, what important outcomes to target and how to go about achieving those desired outcomes. Departing from planned implementation is considered implementation failure. Targeting new and emergent opportunities is considered 'mission drift.' The mantra of traditional, static accountability is plan your work, work the plan, and evaluate whether what was planned was achieved. But that's not how high performance organizations approach either development or accountability.

Henry Mintzberg is one of the world's foremost scholars on strategic thinking, organizational development, and the characteristics of high performing business. He has found that, implementing strategy is always a combination of deliberate and unplanned processes. In studying hundreds of companies over many years, he found that there is no such thing as a perfectly controlled, deliberate process in which intentions lead to formulation of plans, implementation, and the full realization of intended results. The real world does not unfold that way. As the graphic below shows, realized strategy (where you end up after some period of time) begins as intended strategy (planning), but not all of what is intended is realized. Some things get dropped or go undone because planning assumptions proved faulty in the face of real world processes; this he calls "unrealized strategy." What remains of the intended strategy he calls the deliberate strategy, which intersects with emergent strategy to become realized strategy. Emergent strategy comes from seizing new opportunities, which is another reason some things that were planned remian undone as new and better opportunities arise (Mintzberg, 2007, chapter 1). In essence, a high performance organization that is paying attention to the world in which it operates does not expect to rigidly follow a highly prescriptive plan. The plan is a starting point. Once

implementation begins, the plan has to be – and should be – adapated to what is observed and learned, in interaction with the complex adaptive system of real world dyanmics.

Mintzeberg's insights about strategy implementation in the real world contrast significantly with the classic accountability-oriented approach to evaluation in which programme implementation and results are measured and judged based on what was planned to be done and achieved (intended outcomes). Under such an accountability framework, an innovative and adaptive programme that seizes new opportunities and adjusts to changing conditions will be evaluated negatively. Developmental evaluation, in contrast, expects that some of what is planned will go unrealized, some will be implemented roughly as expected, and some new things will emerge. *Developmental evaluation tracks and documents these different aspects of strategic innovation* – and their implications for further innovation and development. Accountability resides in carefully, systematically, and thoroughly documenting these developmental shifts, making transparent the data on which changes are made, and tracking the implications of deviations from the original plan – both deviaitions in implementation and in emergent outcomes.

Figure 2: Mintzberg on Strategy

Complexity-based developmental evaluation shifts the locus and focus of accountability. Accountabilty in developmental evaluation means documenting adaptations and their implications, not evaluating rigid adherence to planned implementation and preconceived outcomes. Why? Because complexity-sensitive developmental evaluation assumes that plans are fallible, based on imperfect information and assumptions that will be proven wrong, and that development occurs in dynamic contexts where even good plans will have to be adapted to changing realities. Thus, rather than becoming a barrier to adapatation, as occurs in traditional rigid accountability

measures in which programmes are deemed to have failed if they depart from what was planned, developmental evaluation assumes a dynamic world with departures from initial plans. Developmental evaluation places the emphasis on understanding, supporting, and documenting adaptations and their implications.

Developmental Evaluation as Utilization-Focused

What brings me to complexity is its utility for understanding certain evaluation challenges. Complexity concepts can be used to identify and frame a set of intervention circumstances that are amenable to a particular situationally-appropriate evaluation response, what I am calling here developmental evaluation. This makes dealing with complexity a defining characteristic of the developmental evaluation niche. Principles for operating in complex adaptive systems inform the practice of developmental evaluation. The controversies and challenges that come with ideas on complexity will also, and inevitably, afflict developmental evaluation. The insights and understandings of complexity thinking that have attracted the attention of, and garnered enthusiasm from, social innovators will also envelope developmental evaluation – and be the source of its utility.

Developmental evaluation is meant to communicate that there is an option in an approach to conducting evaluations that specifically supports *developmental adaptation*. In so doing, I place this approach within the larger context of *utilization-focused evaluation* (Patton, 2008, 2009, 2011b). *Utilization-focused evaluation* is evaluation done for and with specific primary intended users for specific, intended uses. Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration for how everything that is done, from beginning to end, will affect use. 'Use' is about how real people in the real world apply evaluation findings and how they experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on achieving *intended use by intended users*. In developmental evaluation, the intended use is development, which I have here argued is a distinct and important evaluation purpose. The primary intended users are development innovators and others working to bring about major change.

Situation recognition and Developmental Evaluation

Astute situation recognition is at the heart of utilization-focused evaluation. There is no one best way to conduct an evaluation. This insight is critical. The design of a particular evaluation depends on the people involved and their situation. The Development Assistance Committee standards (DAC, 2010) provide overall direction, a foundation of ethical guidance, and a commitment to professional competence and integrity, but there are no absolute rules an evaluator can follow to know exactly what to do with specific users in a particular situation. Recognizing this challenge, situation analysis is one of the "essential competencies for programme evaluators" (CES, 2010)

The ideal is to match the type of evaluation to the situation and needs of the intended users to achieve their intended uses. This means – and I want to emphasize this point – *developmental evaluation is not appropriate for every situation*. Not even close. It will not work if the conditions and relationships are not right. The point here is that every evaluation involves the challenge of matching the evaluation process and approach to the circumstances, resources, timelines, data demands, politics, intended users, and purposes of a particular situation. Such matching requires astute situation recognition. Developmental evaluation is appropriate where the situation is understood to involve interventions and innovations in complex adaptive developmental situations (Patton, 2011a).

References

CES (The Canadian Evaluation Society), (2010). *Competencies for Canadian Evaluation Practice*. http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf

DAC (Development Assistance Committee). (2010). *Quality Standards for Development* Evaluation. http://www.oecd.org/dataoecd/55/0/44798177.pdf

Imas, Linda G.Morra and Ray C. Rist (2009). *The Road to Results: Designing and Conducting Development Evaluations*. Washington, D.C.: The World Bank.

Mathison, Sandra (Ed.). (2005). *Developmental evaluation. Encyclopedia of evaluation*. Thousand Oaks, CA: Sage, p. 116.

Mintzberg, Henry. (2007). *Tracking Strategies*. New York: Oxford University Press.

Patton, Michael Quinn (1994). *Developmental Evaluation. Evaluation Practice* 15, 311-320.

Patton, Michael Quinn (2008). *Utilization-Focused Evaluation, 4th.ed.* Thousand Oaks, CA: Sage.

Patton, Michael Quinn (2009). *Utilization-Focused Evaluation*. In *From Policies to Results: Developing Capacities for Country Monitoring & Evaluation Systems*, edited by Marco Segone, UNICEF, pp. 252-277.

Patton, Michael Quinn (2011a). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* New York: Guilford. http://www.guilford.com/excerpts/patton.pdf

Patton, Michael Quinn (2011b). *Essentials of Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.

Wadsworth, Yoland (2011). *Building in Research and Evaluation: Human Inquiry for (truly) Living Human Systems*. Sydney, Australia: Allen & Unwin.

Wesley, Frances, Zimmerman, Brenda and Patton, Michael Quinn (2006). *Getting to Maybe: How the World Is Changed*. Toronto: Random House Canada.

SYSTEMS THINKING AND EQUITY-FOCUSED EVALUATIONS

Martin Reynolds, The Open University, UK
Bob Williams, Independent consultant, New Zealand

Introduction

Questions about access to resources – who gets what? – ought not to be seen in isolation from related questions of power – who owns what? They also ought not to be seen in isolation from questions of knowledge and expertise – who does what? Moreover these questions relate to important questions regarding legitimacy – who gets affected by what some people get? Such questions are often more easily avoided in a normal evaluation for fear of the ethics and politics involved in addressing them. But such questions as formulated above also may not be easy to grasp or work with in terms of an approach to evaluating an intervention. To the systems thinker C. West Churchman (1913-2004), such ethical and political questions were profoundly important. It was Churchman's life-long task to surface the need to address such questions. One of the most significant insights offered by Churchman in order to address ethical issues was the need to engage meaningfully with different perspectives.

"A systems approach", Churchman famously stated, "begins when first you see the world through the eyes of another" (Churchman, 1968). One of his most influential books "The Systems Approach and its Enemies" (Churchman, 1979), was based on the idea that truly ethical decisions can only be made by considering the perspectives of those whose views you may oppose. Churchman's ideas inspired his doctorate research student, Werner Ulrich, to develop a practical approach to employing such questions in a structured way that could be used by practitioners from different professional and non-professional backgrounds (Ulrich, 1983). Ulrich's approach – Critical Systems Heuristics (CSH) – was a landmark publication in the field of systems thinking. For one distinguished systems practitioner in the field of management sciences, Mike Jackson, CSH "described for the first time an approach that takes as a major concern the need to counter possible unfairness in society, by ensuring that all those affected by decisions have a role in making them"

(Jackson, 2003). This chapter addresses how CSH can be used to explore dimensions of ethics in general and equity in particular, in supporting Equity-focused evaluations. It is illustrated by reference to a case study drawn from rural India.

Equity, evaluation and systems thinking

Equity is an ethic. More precisely, equity is a particular expression of a virtue-based ethic associated with 'justice' that "requires everyone to have the opportunity to access the same resources" (Bamberger and Segone, 2011). UNICEF sponsored evaluations, in short, must ensure good outcomes for redressing the prevailing inequities of resource-access, and/or enhance the rights of access to resources amongst those traditionally worse-off groups in our communities. The Joint Committee on Standards for Educational Evaluation (JCSEE), a touchstone for many evaluators, has an indirect reference to equity:

"Human Rights and Respect: evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders."

Yet evaluation does not have a unified view on who should decide ethical issues relating to an evaluation. Nor beyond the preferences of individual evaluators does it have many established methods of deciding whose perspectives and standpoints should be acknowledged or marginalized. Indeed rarely is it something that an evaluator gets to decide – despite the JCEE standards and similar frameworks adopted by evaluation associations around the world.

The systems field likewise has a strong interest in ethical issues. Systems thinking has gained currency in the evaluation field primarily to assess complicated and complex interventions. The emphasis has been on understanding how multiple factors and actors within situations behave in relation to each other. However, systems thinking also provides a powerful way for evaluators to address important equity issues.

In their book, *Systems Concepts in Action: A Practitioner's Toolkit* (Williams and Hummelbrunner, 2010), Bob Williams and Richard Hummelbrunner argue that systemic approaches to managing inter-ventions can be understood as the confluence of three concepts; interrelationships, perspectives and boundaries. Boundaries deline-ate between what is "in" and what is "out", what is "fair" and what is not. Similarly, Martin Reynolds and Sue Holwell in their book

Systems Approaches to Managing Change: A Practical Guide identify three purposeful orientations for the use of a systems approach in any intervention:

(i) making sense of, or simplifying (in *understanding*) relationships between different entities associated with a complex situation;

(ii) surfacing and engaging (through *practice*) contrasting perspectives, and

(iii) exploring and reconciling (with *responsibility*) power relations, boundary issues and potential conflict amongst different entities and/or perspectives (Reynolds and Holwell, 2010).

The juxtaposition of perspectives, boundaries and power, place systems thinking firmly in the ethical domain, since power assures whose perspective gets to set the boundaries of an endeavour.

Equity and the Narmada Valley Development Project

The Narmada (or Riwa) is the fifth largest river in the Indian subcontinent. It forms the traditional boundary between North India and South India and flows westwards over a length of 1,312 km (815.2 miles) before draining through the Gulf of Cambey (Khambat) into the Arabian Sea, 30 km (18.6 miles) west of Bharuch city in Gujarat.

The Narmada Valley Development Project (hereafter called the Narmada project) in India is not a simple or even a single project.[1] It is better described as a long-term programme involving many individual projects associated with the construction of dams on the Narmada. In systems terms the programme is not merely complicated, involving many variables, but extremely complex, incorporating different and often conflicting perspectives or viewpoints. The Narmada project involves the construction of 30 large, 135 medium and 3000 small dams to exploit the waters of the river and its tributaries for better irrigated agricultural practice to produce more food, and for the generation of hydroelectric power.

---

1   The ideas for the Narmada case study can be found in extracts from Reynolds, M. (2009) "Environmental Ethics" pp. 40-51 in *The Environmental Responsibility Reader*, edited by Martin Reynolds, Chris Blackmore and Mark Smith. London, New York Zed Books and the Open University (downloadable on http://oro.open. ac.uk/18505/1/4ReynoldsEDv2.pdf).

The idea was first conceived in the 1940s by India's first Prime Minister, Jawaharlal Nehru, but it was not until 1979 that the project took form. Of the 30 large dams, Sardar Sarovar is the largest and most controversial. In 1979, the Sardar Sarovar Project was proposed and attracted initial support from international financial institutions including the World Bank. But after much controversy and protest since the late 1980s, particularly regarding the extent of displaced villages and measures to mitigate the extent of such displacement, many financial institutions withdrew support. Protest was led by Narmada Bachao Andolan (NBA), a national coalition movement including people affected by the project, environmental and human rights activists, scientists and academics.

The construction of Sardar Sarovar dam itself was stopped in the mid 1990s. However, in October 2000, the Indian Supreme Court again gave the go-ahead for the construction of the dam. Other dams associated with the wider Narmada project have likewise been the subject of protest.

If we consider "equity" as "fair access to resources" there are four possible evaluation dimensions:

- Water access and quality (e.g. water-borne diseases from stagnant reservoir waters), dealing primarily with water security.

- Urban and rural economic development (e.g. displaced populations from rural areas), dealing primarily with energy security.

- Change in agricultural practices (e.g. shifting towards large-scale irrigated farming), dealing primarily with food security.

- Ecological impacts (e.g. loss of biodiversity in previously rich hydrological systems), dealing primarily with issues of climate change and long-term sustainability.

Issues of equity loom large in any judgement of the worth of the Narmada project. The potential and actual conflicts are clearly formidable.

The Narmada project is far from unique. Large-scale dam construction, like other big socio-economic developments, have been subjected to intense criticism in recent years, both through extensive consultant-reporting and strong activism and protest. But often there is a sense of inevitability about such projects. From a cynical perspective, decisions appear to be made through some inescapable process over which the judgements of evaluators and others

get swept away in the current of so-called progress. But cynicism disguises a wealth of opportunities for seeing and doing things differently; for taking these issues of equity of voice, experience and opinion and doing something beneficial.

A systems outlook on such issues can help to realise such opportunities. For example, some basic systems questions might be asked to reveal areas of responsibility that need to be, and can be, managed more constructively.

- *Interrelationships*. What are the particular issues that need attention and how might they be related with each other? What interrelationships and interdependencies between water, energy and food have a particular impact on worst-off groups? Does global warming deserve more attention than longer standing issues of abject poverty in the world? Or should we just despair at the magnitude and complexity of issues confronting us?

- *Perspectives*. How might these issues be attended to and by whom? Is it just 'them' out there or is it also you/ me/ 'us'? Whose perspectives are relevant to these issues and what realistic role might different stakeholders have in making their perspectives count? How for example may the views of vulnerable groups like pastoralist farmers or other less powerful, and often the most worst-off members of displaced communities such as women, the disabled, and children, be given expression? Or, should we just resort to fatalism, nurturing a general sense of apathy and blame?

- *Boundaries*. Why are some issues privileged more than others, and some ways of dealing with them from particular perspectives prioritised over others? What opportunities are there for challenging mainstream ways of dealing with harmfulness and wrongdoing? Who is and who is not considered an "expert" is a boundary decision? How pervasive are existing systems of expert-driven solutions to poverty-alleviation, or existing systems of financial control by international lending agencies in partnership with national Governments, in sustaining iniquitous situations? Or should we just remain cynical of human nature and the prospects for realising alternative ways of doing things?

Despair, apathy and cynicism are human attributes sometimes encouraged by those with an interest in keeping things as they are – contributing to vicious cycles of business-as-usual and the type of eco-social collapse invoked by cynics. So how might systems

thinking, and more specifically, CSH tools associated with sys-tems thinking, help to overcome such attributes in supporting more meaningful and purposeful pro-equity interventions through pur-poseful equity-focused evaluations?

Systems thinking and Critical Systems Heuristics

The problem is that with so many dimensions to consider, how do you structure a comprehensive equity-based investigation that pre-vents key aspects falling between the cracks and being ignored. Critical Systems Heuristics (CSH) provides a highly structured set of questions that prevents such a fate.

CSH comprises a series of highly evaluative questions designed to expose key perspective and boundary decisions, and subject them to critiques. The intent however, is not simply to expose these deci-sions and subject them to ceremonial ridicule (thereby promoting cynicism) but, in line with Churchman's ideas, to explore equitable ways of resolving the exposed tensions. The twelve CSH questions are associated with four sources of influence – motivation, control, knowledge, and legitimacy – each briefly described below. For a more comprehensive narrative description see Reynolds (2007) and Ulrich and Reynolds (2010).

Motivation

The development of a system – whether it is an intervention itself (e.g., a project, programme or policy) or an evaluation of an inter-vention – starts with some notion of "purpose." Since a purpose reflects embedded values associated with some person or persons, it is valid to ask, "Whose purpose?" Identifying first what the *pur-pose* of the system should be helps identify who the intended *bene-ficiaries* ought to be. This in turn raises questions about what should be appropriate *measures of success* in securing some improvement to those beneficiaries. Together these boundary questions relating to purpose, beneficiaries and relevant measures make transparent the *value basis* of the system.

Control

The exploration of motivation leads to questions regarding the *nec-essary resources* or *components needed for success*. Financial capital and other forms of tangible assets like natural, physical, and human capital might be complemented with less tangible factors

such as social capital (access to networks of influence). But who ought to be the *decision-makers* in control of such resources? This in turn prompts questions as to what should be left *outside* the control of such decision-makers in order to ensure some level of *accountability*. There are risks in having all the necessary resources under the control of the system. If the system has all the resources, then the system cannot be controlled or held accountable in any way by those outside the system. Such questions help make transparent the *power basis* of the system.

Knowledge

One important set of factors that need to be independent of the decision-maker is knowledge or expertise. In an ideal setting, expertise ought not to be under the control of the decision-maker but should have independence. So what ought to be the necessary *types and levels of knowledge and experiential know-how* to ensure that the system actually has practical applicability and works toward its purpose? Who ought to be *the 'experts'*? And how might such expert support prove to be an effective *guarantor*, a provider of some assurance of success? Over-reliance of one area of expertise over the other may constitute a *false guarantee* – a sort of self-deception. Such issues help to make transparent the *knowledge basis* of the system.

Legitimacy

Any assessment of the values (motivation), power (control), and expertise (knowledge) associated with any system will always be biased in some way. So what gives the system the legitimacy to carry out its tasks? If the system is looked at from a different, opposing viewpoint, in what ways might the system's activities be considered coercive or malignant rather than *emancipatory* or benign? Who is capable of *making representations* on the victims' behalf, and on what basis would they make this claim? Finally, how might the underlying *worldview* associated with the system be reconciled with these opposing worldviews? Such questions help to make transparent the worldview or moral meaning underpinning the system. This in turn provides the *basis of legitimacy*; a sense of social and legal approval to the system at any one time.

Figure 1 illustrates the above narrative in terms of twelve boundary questions associated with CSH.

Figure 1: Unfolding narratives of 12 CSH questions

Sources of influence	Social roles (Stakeholders)	Specific concerns (Stakes)	Key problems (Stakeholding issues)
Motivation	1 Beneficiary/client	2 Purpose	3 Measure of improvement
Control	4 Decision maker	5 Resources	6 Decision environment
Knowledge	7 Expert	8 Expertise	9 Guarantor
Legitimacy	10 Witness	11 Emancipation	12 Worldview
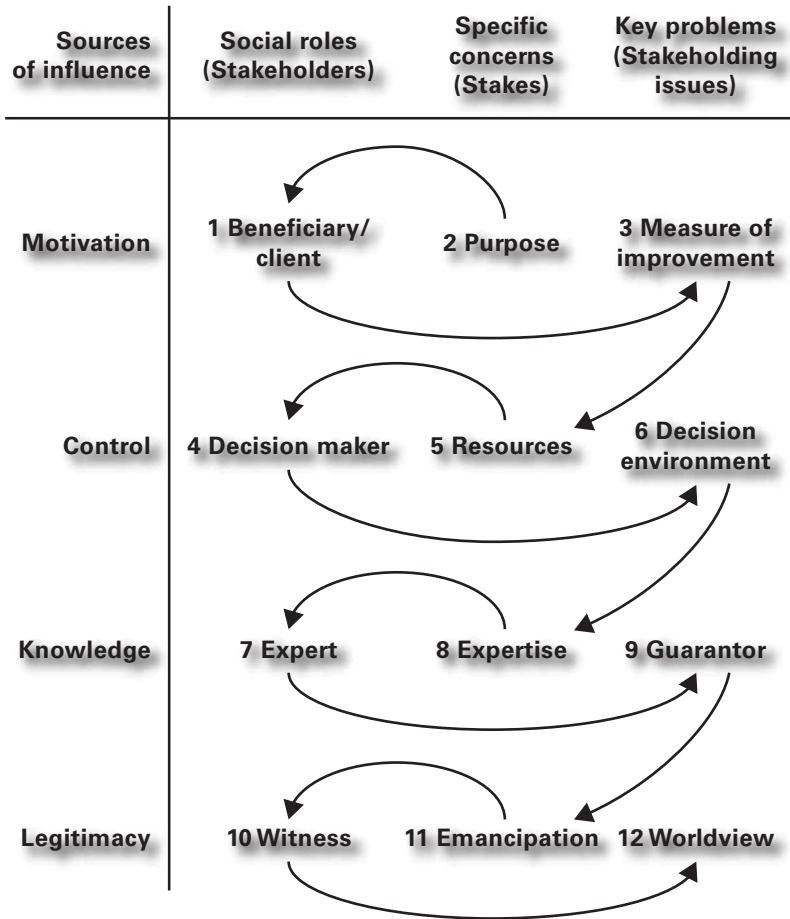
Source: (Ulrich and Reynolds, 2010)

The above narrative covers most of the debates and discussions inherent in a critical systems assessment.

Equity-focused evaluations using CSH comprise three stages:

• firstly, an unfolding of the 12 interrelated boundary judgements in evaluative terms of what 'ought' to be good pro-equity intervention;

• secondly, a critique of the "ought" claims through a more descriptive analysis of the intervention; and

- thirdly, an exploration of possible changes in stakeholding amongst stakeholders in order to improve pro-equity issues being addressed.

The remaining sections examine each stage in detail using the Narmada case study. The Narmada project is not presented here as an actual exhaustive Equity-focussed evaluation study using CSH. The case study is used only for illustrating the three stages and relevance of the CSH questions.

CSH 1: Unfolding key systems questions

You may have noticed that the narrative had an "ought" orientation. In an Equity-focused evaluation it is helpful to have some notion of what a good and/or right pro-equity intervention might look like. Thus the first stage of a CSH inquiry has a normative or "ought" focus. It starts by unpicking the notion of stakes, stakeholders and stakeholdings.

Stakes, stakeholders and stakeholdings

Critical Systems Heuristics starts with a mapping of the interrelationships amongst 12 bounded judgements associated with the issue of stakeholders. CSH makes use of finer distinctions than most stakeholder-based analyses since it starts with three core questions, related to three key boundary decisions:

(i)   What ought to be at stake?

(ii)   Who ought to be the stakeholders?

(iii)   What possibilities ought to exist for improving stakeholdings?

These boundary decisions can be iterative. You normally start by identifying stakes (e.g. access to water; access to sources of energy; access to food; prosperity; soil sustainability; employment; traditional rural life), and on that basis select those who ought to be substantially involved or affected – that is, the stakeholders. The concept of stakeholdings may be new to many evaluators but it is critical to equity issues. With CSH, a stakeholding is a key issue or "problem" related to the topic of interest (say changes in agricultural practice) relating to a particular stakeholder group. So in the case of the Narmada project, one thing at stake from changes in agricultural practice could be traditional rural lifestyles – and from the perspective of a

landlord stakeholder, the stakeholding could be related to the poten-
tially uneconomic nature of patchworks of small landholdings.

If you now review the narrative in the previous section you can
see how each of the four sections were framed in terms of stakes,
stakeholders and stakeholdings (see also Figure 1).

Motivation, control, expertise and legitimacy

The three boundary decisions – who or what ought to be the
stakes, stakeholders, and stakeholdings – are explored from four
distinct perspectives: motivation, control, knowledge and legiti-
macy. These four perspectives are important sources of boundary
critique because:

1. *motivations and values* are built into our view of situations and
   efforts to 'improve' them;

2. *control and power structures* influence what is considered a
   'problem' and what may be done about it;

3. *knowledge* defines what counts as relevant information and
   skills; and

4. *legitimacy* forms the moral basis on which we expect third
   parties (i.e., people not involved, yet in some way concerned) to
   bear with the consequences of what we do, or fail to do, about
   the situations in question.

Stakeholder perspectives between the 'involved' and the 'affected'

Motivations, control and knowledge focus on those *involved* in our
system, legitimacy focuses on those *affected* (often a victim) by
decisions related to motivation, control and knowledge, but who are
not currently part of the system.

Once again, equity here is primarily concerned with issues of jus-
tice. Every system creates 'victims' who have no role or influence
within the system but may well have an influence on the system.
They may be intended or unintended victims. They may be things
(e.g. biodiversity) as well as people. This tension between the
'involved' beneficiaries and the 'affected' victims is signalled in the
last of the CSH questions – the clash of worldviews (Q12).

The CSH matrix

So we now have what amounts to a 3x4 matrix – three stakeholder issues and four sources of influence by which we can critique the system – plus a distinction between 'involved' stakeholders and 'uninvolved but affected' or 'victim' stakeholders. Table 1 provides an overview of these boundary distinctions and describes each of the twelve boundary questions in the normative 'ought' mode. A normative account of a particular system – the Narmada project – is used to exemplify responses to each of the twelve questions.

As Table 1 suggests, CSH exposes different types of stakeholder perspective and potential critiques amongst stakeholder groups. Conflicting interests can be identified between, say, those population groups intended to be beneficiaries of the project (Q1), those with decision-making authority (Q4), and those with relevant expertise to support the project (Q7). Within one stakeholder group conflicts can inform the "stakeholding" column (i.e. Q3, Q6 and Q9).

A higher level critique is between those stakeholders who ought to be 'involved' in the system design (Q1-9), and those stakeholders who ought to be 'affected' by it, but not involved (Q10-12). The tension is particularly expressed in Q12 which signals the space necessary to permit conversation between the worldview underpinning the prevailing system of interest (i.e. aligned with the 'involved') and the worldview of those victims of the proposed intervention/ reference system (aligned with the 'affected').

Table 1: Unfolding normative boundary judgements for an intervention using CSH

Boundary judgements informing a system of interest (S) *For example where S = pro-equity intervention such as the Narmada project*	**Stakeholders**	**Stakes** (Specific interests and motivations)	**Stakeholding issues** (Key problems)
			The involved
Sources of **influence**			
Sources of **motivation**	Q1. *Beneficiary* Who ought to be the intended beneficiary of the system (S)?	Q2. *Purpose* What ought to be the purpose of S?	Q3. *Measure of success* What ought to be the measure of improvement of S?
Narmada project	*Population groups associated with the Narmada River Valley, particularly most vulnerable groups, and Narmada ecosystems*	*Provide wider access to quality water for drinking and agricultural irrigation, hydroelectric energy; whilst mitigating against long-term flooding and ecological damage*	*'impact' measures e.g. general indices of gross national product (GNP), agricultural performance, rural & urban livelihood, poverty alleviation, environmental impact assessments*
Sources of **control**	Q4. *Decision maker* Who ought to be in command of the conditions of success of S?	Q5. *Resources* What conditions of success ought to be under the control of S?	Q6. *Decision environment* What conditions of success ought to be *outside* the control of the decision maker?

Boundary judgements informing a system of interest (S)
For example where S = pro-equity intervention such as the Narmada project

	Stakeholders	Stakes (Specific interests and motivations)	Stakeholding issues (Key problems)
			The involved
Sources of **influence**			
Narmada project	*State representatives of vulnerable poor and the disenfranchised, environmental interest groups, multilateral development banks (MDBs)*	*Capital: finance (including compensation), physical, human (employment), natural (particularly water), social, political (inter-State and international)*	*Federal representatives keeping a check on State representatives; other local, national and international stakeholders and groups not involved in planning; downstream fisheries; forests; wildlife; natural events*
Sources of **knowledge**	Q7. *Expert* Who ought to be providing relevant knowledge and skills for S?	Q8. *Expertise* What ought to constitute relevant-knowledge and skills supporting S?	Q9. *Guarantor* What ought to be regarded as assurances of successful implementation?
Narmada project	*Local rural people, technicians, engineers and scientists with experience in monitoring dam constructions elsewhere, reputable NGO and expert consultancies drawn from local and international contexts*	*Technical: culturally appropriate science & technology, and multi/interdisciplinary skills* *Practical: facilitation skills in developing stakeholder participation, rural peoples knowledge* *Emancipatory: appropriate monitoring of damaging effects and past experiences; transnational expertise*	*Technical: independently guaranteed science & technology, and social science* *Practical: wide stakeholder dialogue/ participation; transnational networking* *Emancipatory: poverty alleviation; triple bottom line guarantors implicit in sustainable development rhetoric of the 1990s*

Boundary judgements informing a system of interest (S)
For example where S = pro-equity intervention such as the Narmada project

Sources of **influence**	Stakeholders	Stakes (Specific interests and motivations)	The 'affected' — Stakeholding issues (Key problems)
Sources of **legitimacy**	Q10. *Witness* Who ought to be representing the interests of those negatively affected but not involved with S?	Q11. *Emancipation* What ought to be the opportunities for the interests of those negatively affected to have expression and freedom from worldview of S?	Q12 *Worldviews* What space ought to be available for reconciling different worldviews regarding S among those involved and affected?
Narmada project	*Advocates of local displaced communities and individuals – the oustee- networked with transnational contacts (e.g. Oxfam); strong representative women's groups*	*Compensation and resettlement of displaced communities and individuals (oustees); alleviation from effects of irregular rainfall and drought; freedom from decisions made in neighbouring States over common resources like water and forests*	*Large scale techno-centric solution good for trickle-down benefits to local communities; economic security of nation and States to be complemented with appropriate concern for existing and predicted skewed socio-economic and ecological development*

CSH 2: A critique of the 'ought' claims and exposure of the 'actual' value base of the system

The normative unfolding of boundary judgements provides a platform for generating an appropriate equity-focused critique of an intervention. So far the assessment has been framed in an 'ought' mode – a value laden *normative* assessment of the situation. It begs the question of whose normative values *actually* dominate. In order to clarify that in a systematic way, the 12 questions are repeated in an 'is' mode – what *is* the purpose (CSHq2)? who *are* the actual beneficiaries (CSHq1)? and so on. The 'is' mode is a *descriptive* assessment of the situation that draws more attention to the actual rather than an espoused value base of the intervention/ system. Each cell is compared using the two modes to generate a set of critiques.

Table 2 provides a summarised critique of the Narmada project from the perspective of each of the four sources of influence.

Table 2: Critique of Narmada project associated with four sources of influence

Sources of Motivation

	Stakeholder (Beneficiary/client)	**Stake** (Purpose)	**Stakeholding** (Measure of improvement)
'ought'	Worst-off Narmada River Valley, groups and ecosystems.	Water, energy, and food security, mitigating against long-term ecological damage.	Poverty alleviation and ecological well-being.

	Stakeholder (Beneficiary/client)	**Stake** (Purpose)	**Stakeholding** (Measure of improvement)
'is'	Less vocal groups like *adivasi* (Scheduled Tribes) prone to being displaced (oustees) and disaffected.	Increase prospect of insect-borne diseases. Inundated areas cause salinization of land alongside canals through build up of salts. Cross-purposes (different aims between different States).	Large numbers of poor and underprivileged communities dispossessed of their land as a source of livelihood; inadequate compensation and rehabilitation for resettled people.
critique 'is' against 'ought'	Is this a case of 'paved with good intentions'? Are multilateral development banks (MDBs) possible surrogate clients?	Is there a 'single bottom line' of national economic development overriding localized socio-ecological development? Unfair and/or unrealistic aims?	Dominance of monetised impacts and indices in terms of GNP masking qualitative impacts ('enchantment of measurable outcomes')? Difficulty in estimating long-term effects and qualitative factors? Possible emphasis on immediate impacts vs. process?

Sources of control

	Stakeholder (Decision-maker)	**Stake** (Resources)	**Stakeholding** (Decision environment)
'ought'	Representatives of vulnerable poor and disenfranchised.	Appropriate capital: financial, physical, human, natural, social, political.	Multiple and appropriate levels of social, economic and ecological accountability.

	Stakeholder (Decision-maker)	**Stake** (Resources)	**Stakeholding** (Decision environment)
'is'	Contestation between 4 State governments sharing common resource. Western multilateral banks still have considerable leverage. Increase power of project's user groups including industrial users of water and electricity.	Long-term dependence on private trans-national companies. Disruption of downstream fisheries. Increased competition over resources by agricultural and industrial users.	Excessive unaccountable profiteering amongst private contractors and possible corruption in dispensing large budgets. Stronger Indian economy diminishes reliance on international funders thereby diminishing sources of accountability.
critique 'is' against 'ought'	Private profiteering from contractors and corrupt officials? Local autonomy vs. Western dependence?	Is this a case of resource development or depletion (e.g. forests & downstream fisheries)? Possibly excessive attention to cash rather than land compensation? What levels of corporate responsibility are evident amongst transnational interests?	Possible command and control ethos in project planning in order to guarantee funding support; lack of accountability for long term effects on displaced communities; or excessive accountability (particularly to MDBs and Federal government) producing delays?

Sources of knowledge

	Stakeholder (Expert)	**Stake** (Expertise)	**Stakeholding** (Guarantor/assurance)
'ought'	Collegiate team of formal and informal (local, including rural peoples) experts conversant with technical and socio-political issues.	Range of technical, interdisciplinary and trans-disciplinary expertise relevant to supporting project aims.	Transparency in levels of certainty and uncertainty associated with project implementation.
'is'	Disparate, wide ranging, and conflicting advisory support.	Loss of traditional local skills in more sustainable farming practices. Reinforcement of, and dependence on, techno-centric expertise, particularly from overseas. Diminish biodiversity through monoculture irrigated farming.	Over-estimate of hydroelectric power generated. False promises regarding maintenance of dams given disorganised State infrastructure and neglect of possible long-term impacts (e.g. large reservoirs could cause earthquakes).
critique 'is' against 'ought'	Over-reliance on experts with vested interests in dam construction, 'green revolution' expertise, and/or indigenous peoples groups associated with international NGOs?	Is there an excessive reliance on international rather than local expertise in reviewing plans? Unexpected consequences generating need for different expertise (e.g. malaria incidence and healthcare).	Increasing and/or changing levels of uncertainty inevitably reduces guarantor provision: reliance on Western 'modern' practices not suited to culture? Risks of earthquakes?

Sources of legitimacy

	Stakeholder (Witness)	**Stake** (Emancipation)	**Stakeholding** (Worldview)
'ought'	Local oustee advocates networked with transnational contacts.	Alleviation from effects of (i) irregular rainfall and drought, and (ii) prevailing top-down planning denying sustainable access to common resources like water and forests.	Platform for expressing equity issues regarding access to resources, intergenerational justice, gender relations, farming practices, energy security etc.
'is'	Proliferation of NGOs in 1980s. Advocate groups like Narmada Bachao Andolan (Ghandian civil resistance techniques) prove effective but divisive between States and less respected with increasing numbers of water-user groups dependent on the project	Evidence of riverine ecosystem damage, along with submergence of forest farmland. Trickle-down benefits to local communities not evident. Transition towards landless rural labour and urbanization.	loss of confidence amongst MDBs due to increased cost factors and strength of protest groups; Narmada project continues despite contestation and micro successes of oustee advocates.
critique 'is' against 'ought'	Conflict between groups representing different affected interest-based constituencies: different effects on same groups situated in different States?	Are costs borne by vulnerable groups being monitored? E.g., control of insect-borne disease, adapting to salinization of soil; ecosystem demise (flooding and deforestation)?	Change in political space for expression by disaffected groups; the project's user groups are too large for project to be abandoned. State and Federal bureaucracies possibly preventing expression of contrasting worldviews?

An equity-focused evaluation of the Narmada project suggests a normative value-base associated with 'equity'. The contrasts between the 'ought' (e.g. in Q1 the intended beneficiaries being prevailing worst-off groups) and the 'is' (e.g. in Q1 the prime beneficiaries are arguably large landowners and those with access to sufficient capital to be able to exploit the opportunities the dam brings) surface concerns about the extent to which the system is actually framed in a way that maximises equity. For instance, some may defend the actual situation arguing that the benefits to large institutions and the capital intensive investments trickle down to the poorest, in ways that many projects aimed at directly alleviating poverty have not done. CSH provides a structure that critiques both stances.

CSH 3: Stakeholding development: exploring opportunities and challenges in the system

It is always good to surface different stakeholder perspectives on, and to provide a critique of, an intervention, not least because it warns against complacency amongst decision-makers associated with an intervention. But, there is always a risk that critiques can lead to an entrenchment of stakeholder positioning or 'stakeholding'.

Stakeholding is a statement of the problem associated with the stakeholder and their stakes. However, it is not the aim of a critical systems analysis to end up with a statement of the problem. The difference between merely stating a problem and actually doing something with the tension it exposes is the difference between 'stakeholding entrenchment' and 'stakeholding development'.

Stakeholding entrenchment is essentially a problem statement that merely reinforces the status quo around a particular intervention. It usually arises from the descriptive part of a CSH critique, and often provides a source of cynicism around stakeholding issues. In contrast, stakeholding development frames these problems as potential opportunities for assisting a deeper resolution of core issues.

Particular attention here is given to CSH questions 3, 6, 9 and 12 regarding stakeholding development.

Key things to look for regarding stakeholding issues

Below is a summary of key issues for an Equity-focused evaluation of the Narmada project.

Sources of motivation ... intended beneficiaries

Q3 Stakeholding (Measure of improvement): *'Key Problems' = tensions between idealised measurable values associated with the performance of the 'system' with a focus on presumed benefits as against realities of a 'situation' with a more balanced view on actual 'costs'.*

Q2 What's at stake? To improve livelihoods of worst-off groups and ecological well-being.

Q1 Who are key stakeholders? People and ecosystems associated with Narmada Valley and their political representatives

Stakeholding entrenchment	*Fixed* 'impact' measures or targets associated with macro-economic 'national' indices e.g. GNP and agricultural performance. Very little attention given to measuring localised impacts and where such measures are used, there is evidence of them being grossly underestimated (e.g. the numbers of project-affected families associated with Sardar Sarovar dam rose from 6.5 thousand in 1979 to over 43 thousand by 2006). Also, despite recommendations in 1969 to compensate oustee (displaced) families with like-entitlements to land and restoration of livelihoods, project officials continued with the tradition of promoting and giving cash compensation (widely regarded as being wholly inadequate and leading to entrenched rural destitution).
Potential for stakeholding development	*Emergent* measures corresponding to changing situations e.g. 1965 report from Narmada Water Resources Development Committee recommending a wider inter-state approach to the initiatives of dam construction taking account of the ecosystem boundaries affected. In 1969 the Narmada Water Disputes Tribunal contested existing colonial Land Acquisition cash compensation laws and replaced them with 'land for land' directives. In the 1980s measures suggested by the Ministry of Environment and Forests (MOEF) required compensatory reforestation, resettlement improvements, wildlife sanctuaries, and other measures arising from environmental impact assessments. More generally though, are there opportunities for measures to be used that may fundamentally question the wisdom of the Narmada project?

Sources of control… decision makers

Q6 Stakeholding (Decision environment): *'Key Problems' =
tensions between idealized control and leverage of the 'system' by national and state bureaucratic institutions as against
realities of a 'situation' with interest groups holding bureaucracies
to account.*

Q5 What's at stake? Command of necessary resources to effect
change in welfare of worst-off communities associated with the
Narmada project.

Q4 Who are key stakeholders? State representatives (primary?);
National government and international representatives (secondary?).

Stakeholding entrenchment	*Command and control* over 'capital' (financial, physical, human, natural, social, political) by corporate industry, multilateral development banks (MDBs) and National and State governments. The entrenched measures of improvement have been circumscribed by regarding the Narmada project as being a fait accompli. The power relationships over such resources have shifted considerably since the Narmada project was first conceived. Numerous 'Committees' have served to consolidate bureaucratic control over the natural resources. Traditionally, land, forests, rivers and fish etc. were communal property managed effectively by local farming practices.
Potential for stakeholding development	*Accountability*: Established checks on State representatives by using National and international stakeholders and groups not involved in planning. A Review Committee was constituted on the back of wider international concern over the plight of displaced communities. Localised groups also emerged providing powerful sources of accountability. For example Narmada Bachao Andolan (NBA) : their non-violent campaigns, including hunger strikes, a 36 day march, mass demonstrations, and use of the media. Friends of River Narmada (FRN): an international coalition of individuals and organizations (primarily of Indian descent) supporting NBA in terms of providing a repository of information, on-going research, public education and outreach, promotion and publicity. The equity oriented evaluation question is how much have these sources influenced and changed existing practices amongst Indian sources of bureaucratic control?

Sources of knowledge… experts

Q9 Stakeholding (Guarantor/ assurance): *'Key Problems' = tensions between idealised promises of the 'system' as promoted by government commissioned experts as against realities of 'situation' through testimony of past experiences and international commentary and cautionary advice.*

Q8 What's at stake? Command of knowledge to guarantee success of the Narmada project as a pro-equity intervention.

Q7 Who are key stakeholders? Rural peoples' knowledge (primary?) National and international government commissioned experts (secondary?).

Stakeholding entrenchment	*Complacent expert (techno-centric) control* over 'knowledge': The Narmada project has all the hallmarks of built-in assurances of success based on a guarantor of good 'faith' rather than appreciating wide-ranging interdisciplinary and trans-disciplinary evidence. Much of the expertise used was international environment consultants employed by the dam construction companies, and therefore lacked independence and trust. Engineers dominated the constitution of the Disputes Tribunal. Considerable evidence existed in India and other developing countries of the underestimated financial and livelihood costs of such projects. These were ignored.
Potential for stakeholding development	*Embracing uncertainty and humility,* acknowledging unpredictability of effects and need for precautionary approach involving wide participation in validation of knowledge deemed relevant. The Narmada Waters Development Committee set up in 1980 used agencies like Caltech and Lokagen (representing rural people's knowledge) and advised on a review of the effects of the dams in 2025, thus acknowledging at least some element of uncertainty. Some movement towards working with independent experts worldwide, at the time when MDBs were under pressure to make their loans more widely accountable towards equitable interests relating to social and ecological responsibility.

Sources of legitimacy... witnesses (advocates)

Q12 Stakeholding (Worldview): *'Key Problems' = tensions between idealised premises underpinning the 'system' dominated by a world-view mantra of 'local pain for national gain' as against realities of a 'situation' in which worst-off groups are further disenfranchised.*

Q11 What's at stake? freedom of expressing disaffection towards the Narmada project.

Q10 Who are key stakeholders? local oustee advocates and environmental groups networked with transnational contacts, and strong representative women's group.

Stakeholding entrenchment	*Relinquishing to pre-dominant assumptions* with inevitability of industrial 'progress' and national economic development through modernisation techniques of agricultural production. The dominant worldview is one that draws its legitimacy from what is known in dam-jargon as the 'iron triangle'. This refers to the insipid relationship of mutual benefits from politics and politicians, bureaucratic entities of corporate control, and construction company expertise. The relationship can be likened to a vicious circle amongst stakeholding interests from sources of motivation (politics), control (bureaucracies), and knowledge (commissioned expertise). The marginalisation of affected groups particularly those indigenous to the areas affected – the *adivasi* – amongst the oustees represents the key expression of entrenchment in Narmada. Adivasi have not been consulted and remain treated as collateral damage to the Narmada project. Despite protests, the project continues to generate 'problems' of landless rural labour demands. They are regarded within project documentation as 'problems' rather than issues to which affected people may have some legitimate role in shaping. Questions arise as to the further denial of political space for deliberating on such issues and the further entrenchment of inequities through pervasive relations of power.
Potential for stakeholding development	*Challenging existing relations of power.* Throughout the 1980s an international network of pressure groups contrived space to lobby MDBs like the World Bank in order to question their support of seemingly iniquitous interventions like the Narmada project. Along with a growing internal awareness of the huge extent of the affects of the Narmada dams on local groups and long term sustainability of riverine ecosystems, this in turn generated space for indigenous NGOs to flourish.

Summary

Critical Systems Heuristics provide a framework by which a very wide range of often complex *interrelationships,* between contrasting *perspectives,* can sometimes lead to uncomfortable and difficult decisions about *boundaries.* Evaluators and their clients who see the world in terms of simple relationships, single perspectives and relatively easy boundary choices will inevitably regard CSH as a challenging process. But they are not the only ones. In their desire to promote a particular world view, or institutional imperative, evaluators, clients and project workers can often ignore the full scope of their endeavour and the issues raised.

Equity-focused evaluations must ensure good outcomes for redressing prevailing inequities of resource-access, and/or enhance the rights of access to resources amongst those traditionally worst-off groups in communities. Often it is difficult to appreciate the wider picture of issues relating to resource access. It is also often challenging to engage with different perspectives about the inequities of resource access.

A systems approach prompts firstly, a greater awareness of the interrelated issues of equity, secondly, an appreciation of different perspectives on inequities, and thirdly a reflection on boundaries used to circumscribe our awareness and appreciation. This third attribute signals the challenges as well as the possibilities of better pro-equity interventions.

Critical Systems Heuristics is helpful in dealing with issues of ethics and politics generally, and hence issues of equity more specifically. The Narmada project provides a particularly rich example of issues of equity addressed with respect to water, energy, and food security, as well as longer term issues of sustainability for future generations.

In using CSH tools to illustrate some preliminary features of an Equity-focused evaluation of the Narmada project intervention, three dimensions of inequities are addressed:

1. Unfolding interrelationships: revealing key equity issues between stakes, stakeholders and stakeholdings; that is, asking:

    (i)   who ought to be getting what (sources of motivation)?

    (ii)   who ought to be owning what (sources of control)?

(iii)   who ought to be doing what (sources of knowledge and expertise)?

(iv)   who ought to be affected by what's going on and how ought any disaffection be given space for expression (sources of legitimacy)?

2. Critique using contrasting perspectives: engaging with issues of possible inequities; that is, asking:

(i)   who *is* getting what in relation to actual benefits?

(ii)   who *is* owning what with regards to control over key resources?

(iii)   who *is* doing what with regards to accepted expert support?

(iv)   who *is* actually affected by what is going on and what actual possibilities are there for such disaffection to be given space for political expression?

3. Stakeholding development: exploring challenges and opportunities of progressing an intervention in accordance with principles of pro-equity intervention, that is:

(i)   reconciling 'fixed' targeted measures of success with pro-equity measures that may be hidden by official 'targets' (are official performance indicators masking existing inequities?);

(ii)   reconciling typical 'command-and-control' decision-making with pro-equity accountability (are decision-makers further marginalising inequities arising from the intervention?);

(iii)   reconciling typical expert promises of project success with inevitable uncertainty and pro-equity cautionary concerns (are experts taking account of the possible inequities arising from the intervention?); and

(iv)   reconciling the premises of an underlying consensual belief system with the need for nurturing contrasting belief systems (are witnesses and advocates of the disaffected effectively opening up – or possibly closing up – space for meaningful and purposeful dialogue?)

CSH is not the only systems approach relevant for equity-focused evaluations. There are many variants of basic systems concepts that can be used for addressing different evaluation questions (Williams and Imam, 2007). However, based on the substantive ethical

principles of systems practice developed by perhaps the most celebrated of systems thinkers, C. West Churchman, and further developed by the philosopher and planner, Werner Ulrich, CSH goes beyond the partiality of bureaucratically convenient 'evaluation questions'. CSH can lay claim to providing one of the most comprehensive and provocative frameworks for evaluating systemic inequalities arising from any intervention.

References

Bamberger and Segone (2011). *How to design and manage Equity-focused evaluations*. New York, UNICEF

Churchman, C. W. (1968). *The Systems Approach.* New York: Dell.

Churchman, C. W. (1979). *The Systems Approach and its Enemies.* New York, Basic Books.

Jackson, M. (2003). *Systems thinking: Creative holism for managers*. Chichester: John Wiley.

Reynolds, M. (2007). "Evaluation based on critical systems heuristics" pp.101-122 in *Using Systems Concepts in Evaluation: An Expert Anthology.* B. Williams and I. Imam. Point Reyes CA, USA Edge Press.

Reynolds, M. (2009) "Environmental Ethics" pp. 40-51 in *The Environmental Responsibility Reader*, edited by Martin Reynolds, Chris Blackmore and Mark Smith. London, New York: Zed Books.

Reynolds, M, and S Holwell. (2010). "Introducing Systems Approaches." pp. 1-23 in *Systems Approaches to Managing Change*, edited by Martin Reynolds and Sue Holwell. London: Springer.

Ulrich, W. (1983). *Critical Heuristics of Social Planning: a new approach to practical philosophy.* Stuttgart (Chichester), Haupt (John Wiley – paperback version).

Ulrich, W., and M. Reynolds. (2010). "Critical Systems Heuristics." pp. 243-292 in *Systems Approaches to Managing Change*, edited by Martin Reynolds and Sue Holwell. London: Springer.

Williams, B & Imam, I. (2007) *Using Systems Concepts in Evaluation: An Expert Anthology.* Point Reyes CA, USA Edge Press:

Williams, B., & Hummelbrunner, R. (2010). *Systems Concepts in Action: A Practitioner's Toolkit.* Stanford, CA: Stanford University Press.

METHODOLOGICAL CHALLENGES IN USING PROGRAMME THEORY TO EVALUATE PRO-POOR AND EQUITY-FOCUSED PROGRAMMES

*Patricia J. Rogers, Royal Melbourne
Institute of Technology University and
Richard Hummelbrunner, Independent consultant*

Introduction

There is increasing agreement that it can be useful for development evaluation to be guided by a programme theory (sometimes referred to as a theory of change or a logic model) – an explicit causal model that articulates how a development intervention, such as a project, programme, policy or strategy, is understood to produce intended outcomes and impacts. However, the potential for programme theory to inform the evaluation of equity-focused programmes, or the equity aspects of programmes more generally, is not always met. This chapter suggests some ways in which this can be done. It begins by discussing some of the key features of equity-focused programmes that programme theory needs to address – in particular, the need to support poor and marginalized people to be agents of their own development, and to address complicated and complex aspects of programmes. It then discusses the implications of these for developing, representing and using programme theory. Programme theory needs to acknowledge the other factors needed to produce intended outcomes and impacts; support appropriate translation of effective interventions to other contexts by distinguishing between theories of change and theories of action; highlight differential effects of interventions, and in particular the distribution of benefits; and support adaptive management of emergent programmes. The chapter includes recommendations for specific changes to traditional forms of programme theory, logframes and results chains, in order to effectively address these changes.

Programme theory: different labels and formats

The idea of basing evaluation on an explicit model of how an intervention is understood to work is not a new one – it dates back at least to Carol Weiss' 1972 book on programme evaluation which sets out a programme theory for a home visiting programme by teachers (Weiss, 1972) and the development of the logical framework approach in the 1970s (Practical Concepts, 1979).

Many different labels are used for this, including causal model, cause map, impact pathway, intervention theory, intervention framework, intervention logic, investment logic, logframe, logic model, logical framework, outcomes hierarchy, outcome line, programme logic, programme theory, programme theory, results chain, theory of change.

All of these refer to an explicit statement of what an intervention is trying to achieve and how. 'How it works' refers to not only what activities are undertaken, but how these are understood to bring about change. For example, if community members are brought together in a group to learn about their rights to services, does it work by increasing their knowledge about their rights, by changing social norms, or by helping them plan strategies to claim their rights?

Programme theory can be represented in different ways, which can be categorized into four broad types: a results chain; a logframe; an outcomes hierarchy; and a realist matrix (Funnell & Rogers, 2011).

A results chain is a common form which represents an intervention in terms of inputs, processes, outputs, outcomes and impact. The following example (figure 1) comes from Canada's International Development Agency (CIDA). This version begins with inputs, which are used to undertake activities, which produce outputs, which lead to immediate outcomes in the short-term, intermediate outcomes in the medium term, and ultimate outcomes in the long term. Some results chains use the word 'impact' for the final level of outcomes in the results chain. This example starts at the bottom and goes to the top, but they can go from top to bottom or from left to right.
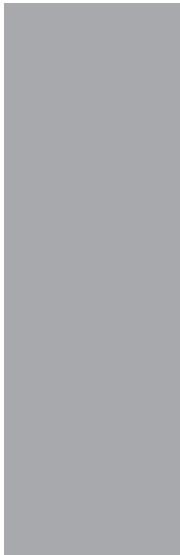
Figure 1: A results chain
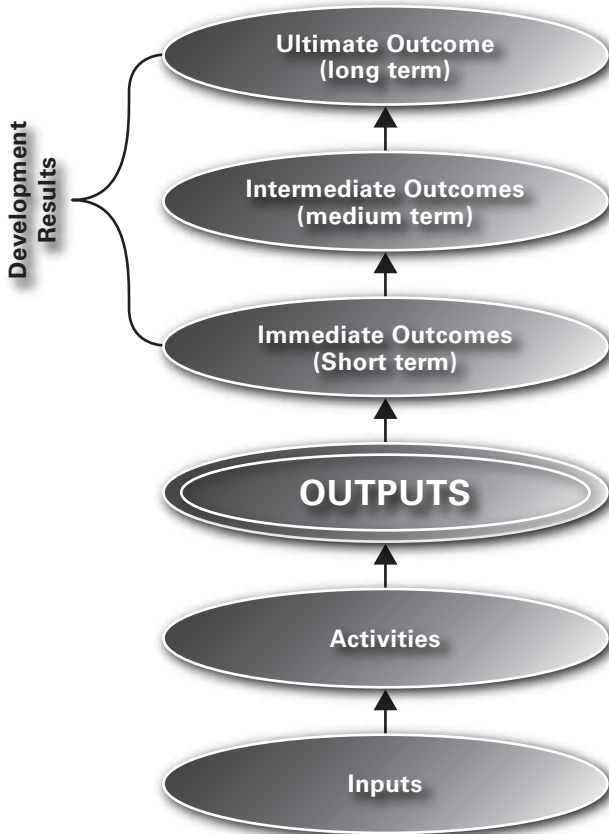
The standardized format of this type of representation makes it easier to generate than the other types, but it has a number of limitations. In particular such representations often fail to explain what it is about the intervention that creates the desired changes, and they assume that all the activities are at the front end of the causal process, rather than occurring throughout the causal process.

A logframe (figure 2) is a particular form of a results chain, developed as part of the logical framework approach, and widely used in international development. The classic version has four components in the causal chain (Activities, Outputs, Purpose and Goal) and for each of these sets out a narrative description, Objectively Verifiable Indicators (OVIs), Means of Verification (MoV) and Assumptions.

Figure 2: A logframe

Project Structure	Objectively Verifiable Indicators	Means of Verification	Important Assumptions
Goals			
Purpose			
Outputs			
Activities			

An outcomes hierarchy (figure 3), sometimes referred to as an outcomes chain or a theory of change, represents programme theory as a series of outcomes from beginning to end. Activities don't have to be up the front, but can be at any stage along the causal chain. This type of programme theory makes it easier to show how different causal strands are understood to combine or conflict in producing the intended results but actual activities need to be shown separately, either in a narrative or a separate programme theory matrix (Funnel, 2000; Funnell & Rogers, 2011).

Figure 3: An outcomes hierarchy
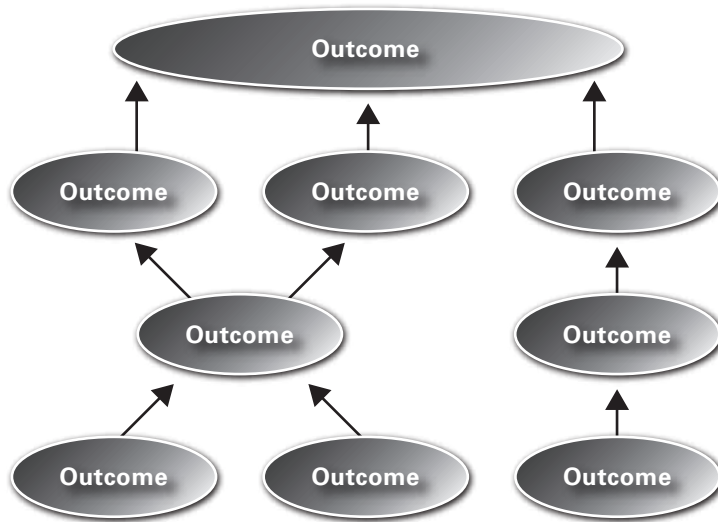
A realist matrix focuses on one causal process at a time and articulates how it is understood to work differently in different contexts (Pawson & Tilley, 1997; Pawson, 2006). It is based on a generative conception of causality, i.e. regarding the outcomes of an intervention as the result of specific mechanisms and contextual conditions. The programme theory is expressed through a set of 'Context-Mechanisms-Outcome' configurations, which serve as assumptions to be tested by evaluations.

Features of equity-focused programmes that need to be addressed

Two major features of equity-focused programmes need to be addressed in evaluations and in the use of programme theory: the need to truly engage and reflect the needs of the poor and marginalized, and to address complicated and complex aspects.

Support the poor and marginalized to be agents of their own development, not passive beneficiaries

One of the important features of equity-focused programmes is that, in addition to improving access to services and other resources, they often include a focus on empowering poor and marginalized

people, supporting them to be agents of their own development. It is important, therefore, that evaluations using programme theory do not inadvertently undercut this strategy and become one more thing that is done <u>to</u> the poor rather than <u>with</u> the poor.

Recognise simple, complicated and complex aspects

Another important feature of equity-focused programmes is that they usually have both complicated and complex aspects.

A number of people have explored the importance of complicated and complex aspects of interventions, including Glouberman & Zimmerman (2002), in their discussion of health services in Canada, and Kurtz & Snowden (2003). More recently, a number of evaluators have discussed the implications of these ideas for evaluation (Patton, 2010; Ramalingam et al, 2008, Forss et al, 2011) and for programme theory in particular (Rogers, 2008, Funnell & Rogers, 2011).

A simple intervention is one where the destination is known, the path to reach it is known, and it is simple to follow the plan. Furthermore 'one size fits all' – its programme theory provides a recipe for success that can be readily transferred to different situations. The intervention is sufficient within itself without other contributing factors, such as the influences of other programmes or favourable implementation environments. Few development interventions totally fit this description, however programme theory often represents them as if this were the case.

Most development interventions have complicated aspects. They have multiple components which are needed to produce the intended results. It is possible to plan in detail the implementation of an intervention, but it requires considerable expertise and coordination of different components. Programmes work differently in different situations. The programme theory from one intervention can be adapted to suit a different situation.

Many development interventions also have complex aspects. 'Complex' does not mean 'very complicated' but refers to emergent, adaptive interventions where it is not possible to set out in advance the details of what will be done, even with expert input. In a complex situation, it is not possible to create a detailed plan and stick to it, because the situation is changing rapidly, and the components are changing in response to this and to each other. Many development interventions work largely through creating enabling conditions, and then applying adaptive management to address emergent issues, dampening negative influences and amplifying positive ones.

A single development intervention might well have some simple aspects, some complicated aspects, and some complex aspects. A mapping project from Ushadishi that produced a map of incidents and resources in post-disaster Haiti illustrates this. It has complicated aspects – for example, there are different levels of activity, and different components needed for success. It also has complex aspects – for example, these maps were jointly created by decentralized, independent people, who interact together without centralised control, identifying sites with needs and also resources that could meet those needs.

Therefore, for programme theory to be helpful in guiding evaluation it is likely to be more useful to address complicated and complex aspects, rather than representing the whole programme as if it were completely simple.

Implications for developing programme theory

Programme theory can be developed when a programme is being planned, or when an evaluation is being planned, and often needs to be reviewed and revised during implementation. Programme theory can be developed from formal programme documents such as stated goals, published research on similar programmes, observations of the programme in action, tacit knowledge of programme staff, and mental models of staff, community members and other key stakeholders. Ideally it is developed through a combination of sources.

Involve different stakeholders in developing programme theory, including the poor and marginalized

Programmes that aim to improve equity need to be based on a good understanding of the perspectives of the poor and marginalized. Participatory approaches to monitoring and evaluation are based on the premise that the poor are active participants in development interventions, and their actions should be informed by active participation in monitoring and evaluation. This means being involved in developing the programme theory (clarifying the intended impacts of the intervention, and how these are intended to be achieved) and undertaking the evaluation (not just providing data, but also deciding what the evaluation needs to focus on, and what would constitute credible and useful evidence).

Transformative evaluation (Mertens, 2008, and her chapter in this volume) specifically focuses on ensuring that the process of evaluation engages those who are intended to benefit from an intervention. Participatory approaches to developing and using programme theory are appropriate in these situations. The same holds true for the use of systems concepts and methods (Williams & Hummelbrunner, 2011). Of particular relevance are those approaches that deal with the capturing of various perspectives and a critical reflection of boundary choices (see both Williams and Reynolds in this volume).

Beneficiary Assessments (BA) and Participatory Poverty Assessments (PPA) explicitly involve the intended beneficiaries in the process of undertaking a situation analysis, which is the first stage of developing a programme theory. These approaches use a combination of conversational interviews, focus groups and observations to explore questions such as those set out in table 1.

Table 1: Questions for developing a Programme theory

Popular perceptions of problems

- How do the poor perceive various manifestations of poverty? This would include income factors, work opportunities and conditions, nutrition/food security, vulnerability to drought, natural disasters, and violence.

- What do they see as the root causes of poverty? This might include war/conflict, exploitation, insufficient assets, or access to services, gender inequalities, sexual or ethnic discrimination, and lack of representation.

Incentive and Regulatory Framework

- What are the micro and macro level factors that affect the accumulation of human capital and access to land and credit as the poor see it in their country?

- How seriously are persons of various ages and ethnic and racial groups affected by imperfections and inequalities in the labour market?

- How are diverse groups of the poor affected by price increases of key commodities?

- What changes in the incentive system would be most welcomed by the poor?

Source: World Bank (2002) Beneficiary Assessment pp5-6.

Beneficiary Assessments (BA) can be useful in terms of developing programme theory in three ways. Firstly, they can identify what outcomes are valued among the poor. For example, a BA undertaken in Mali established that the reason behind low enrolments in primary school was that parents did not see that the benefits were worth the costs (World Bank, 2002). Secondly, BAs can identify what is

needed to engage people in the intervention (the 'invisible mecha-nisms' discussed by Pawson (2009). For example, a BA in Lesotho found that a health service had failed to engage people because of its exclusive emphasis on preventative health, rather than providing some direct curative treatments (World Bank, 2002). Thirdly, BAs can provide information about which theories of action are likely to be effective in triggering theories of change in different contexts. For example, a BA in India had found that radio was the best way of communicating information about maternal and child health. How-ever, a BA in Madagascar found that radios were unsuitable ways to provide information to farmers due to the low level of access to functioning radios and batteries. Instead oral exchanges in the mar-ketplace were the way this information would be accessed (World Bank, 2002).

Focus on understanding how change comes about, not just what happens

Programme theory needs to go beyond just listing activities and intermediate outcomes (connected by arrows which do not actually explain how one leads to the other) and articulate an actual theory of change – how is it understood that change comes about. Since equity-focused programmes are often designed to bring about change at a number of levels – individuals, households, organi-zations and communities – each of these might need a theory of change.

It can be helpful to draw on previously identified theories of change and programme archetypes if they are relevant. For example pro-grammes can work by providing information to inform decisions, and thereby change incentives (carrots and/or sticks); case man-agement; community capacity building; and direct service delivery (Funnell & Rogers, 2011).

Within development, there are a number of classic theories of change and programme archetypes which have associated pro-gramme strategies (table 2).

Table 2: Archetypes for Change in Development

Archetype: How Change Happens	Change Strategy: What we do
Active Citizenship: four powers; integrated change strategy using multiple strategies	Active Citizenship: people in the streets; popular mobilization; supporting grassroots organization
Active Citizenship: grassroots leadership	Leadership training
Elites: enlightened leaders	Advocacy and elite networking
Elites: technocrats make evidence-based policy	Research-based advocacy
Cross-Class: democracy works	Election campaigns, party influencing, voter registration drives
Cross-Class: coalitions of dissimilar players (e.g. civil society, private sector, sympathetic state officials) drive 'transitions to accountability'	Alliances and coalitions; convening role; use of power analysis to design insider-outsider advocacy and programme strategies
Dynamics: steady incremental progress	Focus on binding constraints
Dynamics: tipping points and breakthroughs	Reactive: rapid shift of resources to respond to shocks (financial crisis, Arab Spring etc)
Dynamics: contagion, through the power of example	Piloting/supporting new approaches, publicising success
Dynamics: non linear and evolutionary	Accelerating evolution: supporting experiments, helping with variation and selection; advocacy for amplification

Source: from Green, 2011

Within specific types of programmes there are other specific theories of change. For example, table 3 gives some of the theories of change identified as underpinning peace-building programmes, all of which could also be relevant for equity-focused programmes.

Table 3: Theories of change for peace-building
programmes

Theory of change	Related programme activities
Individual change: transformative change of a critical mass of individuals	Investment in individual change through training, personal transformation/ consciousness-raising workshops or processes; dialogues and encounter groups; trauma healing
Healthy relationships and connections: break down isolation, polarization, division, prejudice and stereotypes between/among groups	Processes of inter-group dialog; networking; relationship-building processes; joint efforts and practical programmes on substantive problems
Root causes/justice: address underlying issues of injustice, oppression/exploitation, threats to identify and security and people's sense of injury/victimization	Long term campaigns for social and structural change, truth and reconciliation, changes in social institutions, laws, regulations and economic systems
Institutional development: establish stable/ reliable social institutions that guarantee democracy, equity, justice and fair allocation of resources	New constitutional and governance arrangements/entities; development of human rights, rule of law, anti-corruption; establishment of democratic/equitable economic structures; decentralization
Grass roots mobilization: mobilizing community so politicians have to pay attention	Mobilize grass roots groups, non-violent direct action campaigns, use of the media, education/mobilization efforts, advocacy groups

Source: Church and Rogers (2006)

Understand the intervention as embedded in its environment

Equity issues rarely can be addressed by a single intervention. It is therefore important to conceive the intervention as part of a greater whole and identify the factors that connect the two. Programme theory needs to identify the contributions from other programmes or contextual conditions that are needed if interventions are to achieve equitable outcomes and impacts.

For example, conditional cash transfer programmes might be effective in increasing school attendance but will only produce the intended impact of improved learning outcomes if the school

system is functioning well enough to educate students who attend. Programmes to improve teacher performance by increasing teacher salaries will have different effects in situations where the cost of these additional payments comes from a fixed education budget, leading to other negative effects (such as larger class sizes, or reduced funding for textbooks). In agricultural development, there is now increasing attention to the need to develop value chains to achieve intended poverty reduction impacts – simply producing more agricultural products is not enough without also developing access to markets for selling these.

Identify distributional effects – not just mean effect

Many interventions have heterogeneous impacts –producing positive results for some people, having little effect for others, and even having a negative impact on others. Focusing only on the mean impact hides these differential impacts. At its worst, this can lead to programmes being supported that are actually harmful for the most vulnerable. For example, evaluations of early childhood programmes, have found that many programmes, while having on average, a positive impact, had negative impacts (that is it made things worse) for the most vulnerable families, presumably the ones the programme was most intended to support (Westhorp, 2009).

This presents quite a challenge for programme theory which rarely shows different causal paths for different groups. Realist matrices are a way to identify and understand differential impacts across different types of participants and context conditions identifying "what works for whom under which conditions".

Implications for representing programme theory

Choose appropriate ways of representing programme theory

A programme theory is often presented in a concise form (e.g. graphically) to facilitate joint understanding and communication among the various actors involved. If the programme theory contains complicated and/or complex aspects (see above), the form of representation should allow for capturing them.

In the case of *complicated aspects* this could mean to:

- elaborate more refined graphic representations that show the interrelations between multiple components or allow the capturing of various causal strands (sequential and/or in parallel)

- show the intended linkages with other interventions or with contextual factors that are considered important for the success of an intervention.

Figure 1: Using a logic model to represent a programme theory of an equity-focused intervention

The model can either be used for a stand-alone equity-focused programme or to assess the specific equity-effects of incorporating an equity strategy into a conventional service delivery programme
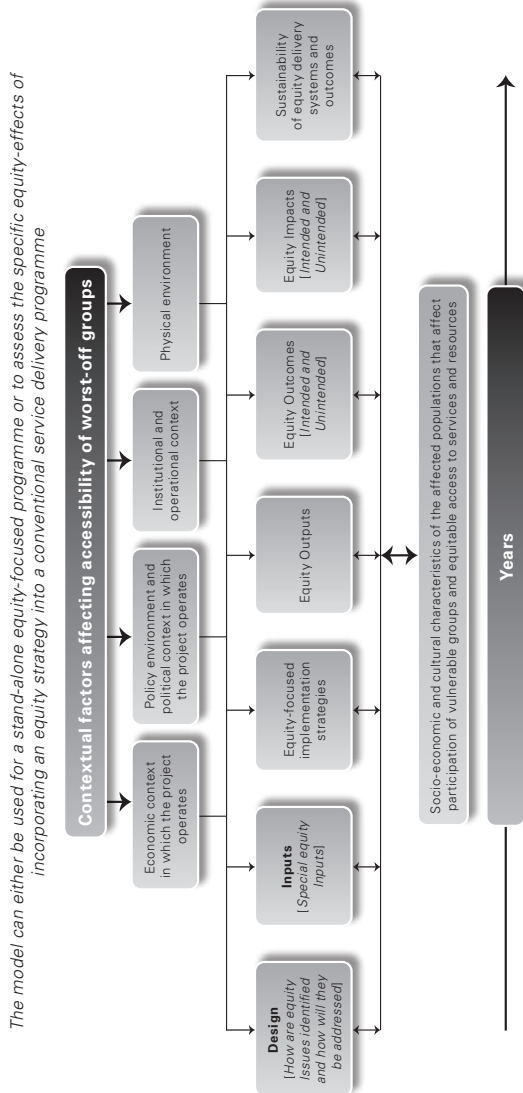
Source: Bamberger M. and Segone M. (2011)

Some *complex aspects* can also be represented via logic models (e.g. mutual or reciprocal influences between elements), but for capturing non-linear relationships, different forms of representation will be needed:

• Feedback mechanisms are best illustrated through Causal Loop Diagrams, which visualize variables and their relationships over time. These are expressed as 'positive' (reinforcing) or 'negative' (balancing) feedback-loops, and their combination allows analysis of relational patterns and mapping of the structure that is responsible for producing recurring events over time.

• Tipping points (i.e. changes in behaviour beyond a certain threshold) are best captured through stock-and-flow diagrams used in System Dynamics. The variables in these diagrams must be quantified and their relations defined as mathematical functions. This allows a simulation model to be built to test and explore the dynamic behaviour of a situation.

Another possibility is to choose representations that permit the *highlighting of specific features* considered to be important. Figure 2 shows the programme theory representation of a small-scale agricultural project, which emphasizes (and categorizes) the programme`s impacts on the lives of the poor. Moreover, since this programme involves testing and adapting technologies, this part of the theory of change is represented as a learning cycle that connects the corresponding activities and outcomes.

Figure 2: Representation highlighting specific features
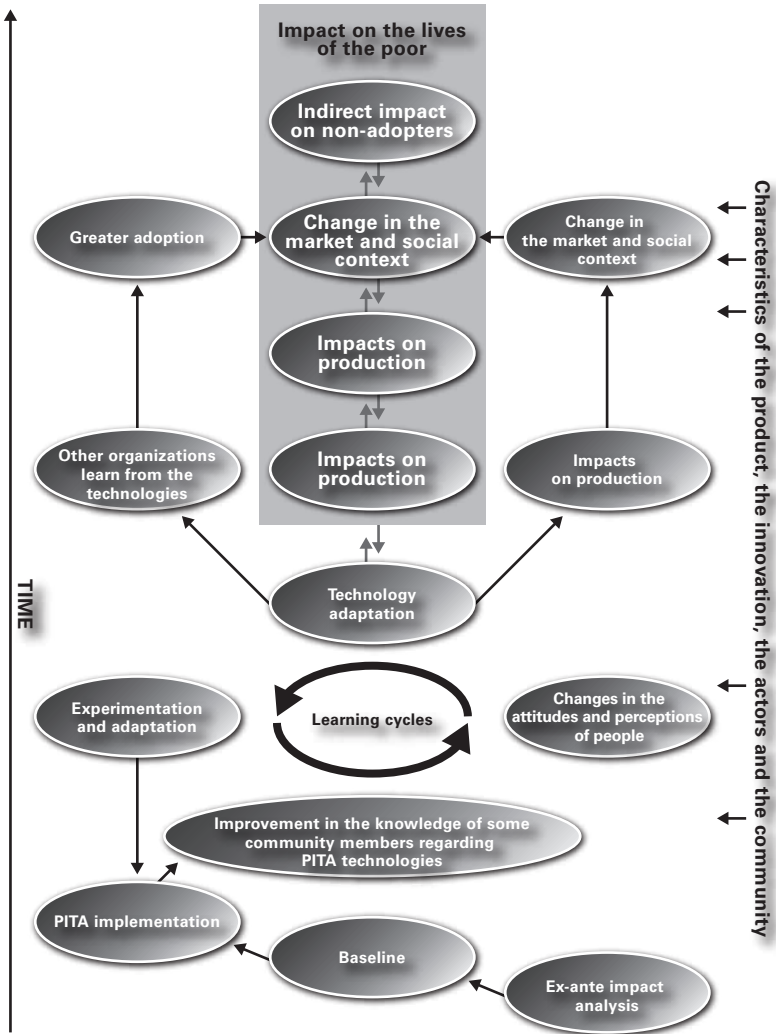
Source: Paz R., Dorward A., Douthwaite B. (2006).

Due to their format (matrix with narrative descriptions), Logframes offer fewer possibilities for refinement. Moreover, they assume a linear and quasi-automatic progression of effects, by which carrying out activities as planned is a guarantee for the achievement of expected outputs or purposes. This tends to ignore all the other factors that might be at play, in particular in complicated and complex situations.

To render Logframes into a more useful tool under such conditions, *situational recognition* should be incorporated by differentiating relevant Logframe components. For instance, the expected outputs can be categorized into the three domains (simple, complicated, complex) based on the (dis)agreement/ (un)certainty parameters advocated by Glouberman and Zimmerman. The outputs can then be clustered by using the portfolio technique, as illustrated in figure 3.

Figure 3: Situational recognition
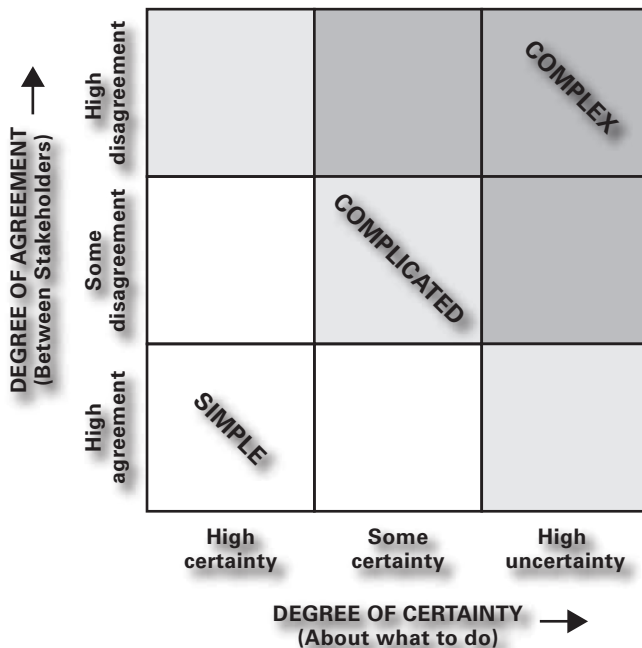
If considered more appropriate, this categorization can be done at the level of activities or outcomes. Such a differentiation has implications for completing other Logframe elements and for the utility of Logframes altogether:

- If most of the outputs fall under the 'simple' domain the Logframe can be confidently used as a monitoring tool.

- If outputs predominantly lie in the 'complicated' domain, indicators and assumptions should be carefully identified to enable monitoring of effective practice, relevant factors and context conditions. Complementary approaches (e.g. logic models) should be used alongside Logframes to obtain a sound logic of intervention.

- If many (or even the majority of) outputs are considered to be 'complex', the indicators should allow for documenting initial conditions and – in combination with assumptions – capture emerging phenomena. But even with these precautions, Logframes will not be a reliable monitoring tool as the programme theory needs to account for emerging phenomena or feedback loops between elements. These features cannot be represented in Logframes and other methods based on systems thinking or complexity theory must be applied, which are suited to identifying relevant patterns between variables (or actors).

Making these adjustments to the content and use of Logframes can mitigate some of their conceptual limitations as a planning and monitoring tool, acknowledging that some of the Logframe´s underlying beliefs – perfect advance knowledge, full control of implementation – are not valid in all situations. And it will help to overcome the inherent tendency of Logframes to treat all development interventions (in their entirety) as 'simple'.

Equity-focused interventions often achieve their objectives through the promotion of behavioural changes and contributions by specific actors or by ensuring that expected effects reach specific target groups. In such cases the *actor dimension* should be captured when representing the programme theory. This can be done in several ways:

- An influence matrix can be used to show the intended linkages between effects (e.g. outputs) and actors (either as contributor or as beneficiary), which also allows capturing multiple relationships.

- Each level of a Logframe can be associated with a set of actors, who are expected to collaborate for their achievement. Their relationships can be shown at each level and also across levels (e.g. by using Social Network Analysis). Thus the time stages are complemented by a sequence of actors, which can be seen as pathways along which information, resources and material objects can pass – in both directions. Such a Social Framework view allows the illustration of the distribution of responsibility for achieving effects across a wider network of actors, and the description of the respective changes expected from each actor (Davies, 2005).

- For interventions involving social change processes or where capacity building plays a major role, Outcome Mapping is a suitable way to articulate the underlying programme theory (Earl et al, 2001). The focus is on one specific type of outcome, i.e. changes in the behaviour of relationships, actions or activities of the people, groups, and organizations with whom an external agent is working directly and seeking to influence. Lately, some attempts are being made towards a fusion of the Logframe and Outcome Mapping. Ambrose and Roduner (2009) postulate that such a fusion can integrate the results-oriented focus of Logframes with Outcome Mapping´s process-oriented learning pathways. Figure 4 shows how elements of Outcome Mapping (e.g. Outcome Challenges, Progress Markers) can be inserted in the Logframe structure.

Figure 4: Inserting elements of Outcome Mapping into Logframe

Formulation of the desired
situation as well as
the practices and behaviour
of project partners

**Overall Goal
(Beneficiaries)**

Impact
indicators

Description of concrete
changes to be attained
by the project
(situational changes)

Programme Goal

Goal
indicators

Description of tasks and
responsibilities of
programme partners

Programme Partner A
Outcome Challenge

Programme Partner A
Outcome Challenge

Progress
Markers
(Outcomes
Measurement)

Definition of tasks and
roles of the project team
& outputs that are
provided to partners

Output

Output

Output
indicators

Strategy map

Strategy map

Mission: Defining
the intended
overall support by
the external
change agent

Mission of the Change Agent &
Project Management
(Budget, HR, Organisational Practices)

Capture the views and perspectives of different stakeholders

First of all, this means departing from the notion that a programme theory requires identifying – and reaching consensus on – one logical path from activities to goals. In practice, such a single logic is rather the exception and differences in opinion among stakeholders concerning strategy or effects often prevail. Yet, many representations (Logframes in particular) do not lend themselves to incorporating these differences, or taking into account various perspectives. As a result, they often fail to reflect the messy realities facing development actors, thus producing confusion rather than clarity.

There are several ways this limitation can be overcome (provided the programme theory is developed in a participatory manner): one

could elaborate separate graphic representations (for specific stakeholder groups or clusters of stakeholders) or integrate relevant differences in opinion into a consolidated picture. Different perspectives can also be described verbally, and such a document can be used to complement a graphic representation.

The latter can also be done with Logframes: capturing diverging views on Logframe elements in a separate document would not only permit articulation of views on their importance, but also address their distributional effects for different actors (in particular, the poor). Another possibility is to capture the opinions of stakeholders on the appropriateness and pro-poor orientation of certain effects (e.g. Outputs, Purpose). Using the same portfolio technique outlined above would permit the assessment of the range of (dis) agreement between stakeholders on the appropriateness of outputs as well as on their degree of certainty with respect to pro-poor effects.

Incorporate assumptions in line with the situation

A programme theory should make assumptions about internal and external conditions, which can support or hinder the achievement of intended effects. If assumptions are used for monitoring purposes, i.e. to track progress towards achieving effects, they should focus on those factors that can be influenced or reacted upon by the intervention (e.g. through one's own mechanisms or via partners). In this way they articulate uncertainties and describe risks associated with an intervention. Identifying and monitoring these risks will be particularly important for aspects that fall under the complicated and complex domain.

Assumptions are often formulated in vague or general terms, which limit their utility for monitoring. Clear articulation and localisation of assumptions is therefore a prerequisite for spotting relevant developments during an intervention`s implementation.

It is important to bear in mind that due to their focus, different representations allow for inserting and localizing assumptions in a different manner:

- Logic Models: in general neither theory of change nor programme logic models include assumptions. But these can be inserted (e.g. as a bulleted list) along the entire causal chain and are often informed by preceding knowledge about how an intervention should work (based on theory, experience, research, values etc.).

- Outcome Hierarchy: assumptions can be formulated for the achievement of each outcome (also in combination with required activities) and can be localised all along the causal chain.

- Outcome Mapping: assumptions are made about the behaviour changes of key ('boundary') partners that are needed to achieve intended effects ('Outcome challenges'). And indicators are identified to assess progress in making these changes ('Progress Markers').

- Realist Matrix: assumptions are made for the linkages between outcomes and their key causal variables, differentiating between internal (mechanisms) and contextual factors. These assumptions are expressed as 'Context-Mechanism-Outcome Configurations'.

These options and possibilities should be borne in mind when deliberating the alternatives for representing programme theory.

Despite its conceptual importance, the Assumptions & Risks column is often the Logframe element that is taken the least seriously, filled in at the last minute and in rather general terms. Also it is frequently forgotten that this column is the core feature for expressing a Logframe's vertical logic, connecting the various levels in a Logframe. It should therefore address the processes, which are expected to transform the achievements of one level (e.g. activities) into the next level of effects (e.g. outputs). Assumptions should particularly describe expected (behaviour) changes of specific actors, contributing factors from other interventions or relevant contextual factors. They can also articulate intended combinations between Logframe elements (e.g. various outputs expected to contribute to a purpose) or between the intervention and external/contextual factors.

Implications for using programme theory

Develop appropriate indicators

An indicator is a variable for measuring achievements or to reflect changes connected to an intervention. In result-based management, and with respect to the accountability discussion, the notion of *SMART* indicators has become widespread in international development:

S Specific;

M Measurable;

A Achievable;

R Relevant;

T Time-bound.

Developing SMART indicators means that the variables should be completed with a timeframe, a baseline and target values. Consequently, these indicators are more specific; they might also include information about target groups and what needs to be achieved for these target groups, which is of particular importance for equity focused evaluations.

Indicators should also be chosen *in line with the characteristics* of an intervention:

• In the case of complicated aspects, indicators should enable monitoring of effective practice, relevant factors and context conditions.

• For complex aspects, indicators should allow for documenting initial conditions and – in combination with assumptions – capture emerging phenomena.

An aspect that is often neglected is the *time dimension* of indicators. With respect to their timing in relation to an effect (e.g. output), indicators can be classified into three categories:

• Leading indicators provide information before the corresponding effect takes place, thus they can be considered 'early-warning signals'. Such indicators often relate to qualitative aspects that lend themselves for verification rather than measurement.

• Coincident indicators show a value at about the same time as the effect actually takes place, thereby providing information about the current state.

• Lagging indicators provide data only after the effect takes place, often with considerable time lag, which can be due to (statistical) data collection routines or long result chains.

As a rule of thumb, Leading and Coincident indicators are best suited for monitoring, whereas Lagging indicators should be used for final reporting and evaluations, since data will most likely only be available at the end or even after an intervention's implementation period. For 'Complicated' aspects Leading indicators should be used as much as possible, which enable the monitoring (and review) of unfolding practice as well as the tracking of progress towards achieving outputs and outcomes. In the case of 'Complex' aspects, Leading indicators should be chosen that are sensitive to

small changes and for which baseline values can be established in order to document initial conditions.

Look for differential effects

Many interventions have heterogenous impacts – being positive for some people, having little effect for others, and even having a negative impact on others. Focusing only on the mean impact hides these differences, which might be crucial for equity-focused inter-ventions. If differential effects are neglected, this can – at its worst – lead to programmes being supported that are actually harmful for the most vulnerable. For example, the average impact of an early childhood programme might be positive, but it can have negative impacts (making things worse) for the more vulnerable families – which might be the ones the programme actually intends to sup-port.

Identifying distributional effects presents quite a challenge for pro-gramme theory and their representations. There are several options for *capturing heterogenous impacts*:

• Representations of programme logic that show different causal paths for different groups.

• Realist matrices that allow identification and understanding of differential impacts across different types of participants and context conditions, thus identifying 'what works for whom under which conditions'.

• Programme theory matrices, which articulate 'what success looks like', e.g. by stating intended outcomes and impacts in terms of their impact for the poor.

Another way to capture differential effects is *disaggregating data* according to relevant equity categories, to ensure that information is gathered separately for specific target groups. If sex is a relevant category this would meant differentiation e.g. the primary enrol-ment rate (boys/girls), the number of health professionals trained (by male/female) or the use of latrines (men/women). The same holds true for choosing indicators, which should be those that are sensitive for equity issues. If gender is an important issue, this could mean selecting indicators such as the number of women teachers working in remote areas, the number of new training places open to women, or the poverty rate in female-headed households.

To ensure that equity issues are dealt with across the entire pro-gramme theory, the representation of equity can be scrutinized

from a specific perspective. An example is the *engendering of a Logframe*, which requires that the process of planning, as well as the components of the logframe matrix, is seen through a 'gender lens.' This lens is informed by gender analysis, which is a methodology to investigate the socially constructed differences between men and women, and between women themselves. The preparation of an engendered logical framework matrix involves project planners, stakeholders, and beneficiaries in analyzing gender relations and addressing questions at each level of the framework (see table 5). This analysis should not only be undertaken once during start-up, but be up-dated throughout the course of monitoring and evaluation (Hambly Odame, 2001).

Table 5: Questions for 'Engendering' a Logframe

	Indicator	**Sources of Verification**	**Assumptions/ risks**
Overall goal Do gender relations in any way influence the project goal?	What measures can verify achievement of the gender-responsive goal?	Are the data for verifying the goal sex-disaggregated and analyzed in terms of gender? What gender analysis tools will be used (e.g. in impact assessment)?	What are the important external factors necessary for sustaining the gender-responsive goal?
Project objective Does the project have gender-responsive objectives?	What measures can verify achievement of the gender-responsive objective?	Are the data for verifying the project purpose sex-disaggregated and analyzed in terms of gender? What gender analysis tools will be used (e.g., in Rapid Rural Appraisal exercises)?	What are the important external factors necessary for sustaining the gender-responsive objective?

	Indicator	Sources of Verification	Assumptions/ risks
Expected results Is the distribution of benefits taking gender roles and relations into account?	What measures can verify whether project benefits accrue to women as well as men, and the different types of women engaged in or affected by the project?	Are the data for verifying project outputs sex-disaggregated and analyzed in terms of gender? What gender analysis tools will be used (e.g. in participatory field evaluations)?	What are the important external factors necessary for achieving project benefits (specifically, benefits for women)?
Activities Are gender issues clarified in the implementation of the project (e.g. in workplans)?	Inputs: What goods and services do project beneficiaries contribute to the project? Are contributions from women as well as men accounted for? Are external inputs accounting for women's access to and control over these resources?	Are the data for verifying project activities sex-disaggregated and analyzed in terms of gender? What gender analysis tools will be used (e.g. in monitoring the activities)?	What are the important external factors necessary for achieving the activities, and especially for ensuring the continued engagement of male/female participants in the project?

Source: Hambly Odame H., (2001).

Support knowledge translation as well as knowledge transfer

Evidence-based policy and practice is not just about finding out 'what works' and then doing it. It necessarily involves a process of generalizing from one situation to another. Developing programme theories which articulate how contextual variables (implementation environment and participant characteristics) influence the achievement of intended impacts can inform the appropriate use of evidence, helping to identify situations that are similar enough in the ways that matter for replication to be effective.

It can be useful to distinguish between knowledge transfer (when knowledge about 'what works' can be used in a new setting without making any changes to the intervention) and knowledge translation (when this knowledge has to be adapted to suit the new situation).

Knowledge translation often involves making changes to the way an intervention is implemented (its 'theory of action'), while keeping the underlying logic (its 'theory of change') the same. For example, a successful intervention might be based on increasing knowledge of participants through knowledge dissemination by a credible informant. In one setting this might involve written material produced by a respected local organization; in another setting this might involve oral briefings by peer tutors. While the implementation would look very different, both cases are based on the same underlying theory of how change comes about.

Adapt the programme theory as needed

Programme theories should be dynamic, subject to change throughout the life of the intervention to which they refer. Otherwise they risk being out of touch with the implementing reality and cannot fulfil their function to guide monitoring and evaluation as well as adapting to emerging challenges and opportunities. So programme theory is not a blueprint to be developed in advance and then followed, but a heuristic that needs to be reviewed, revised and refined as implementation proceeds. This holds particularly true for interventions that have complex aspects, which cannot foresee or specify everything in advance.

In theory, many of the formats for representing programme theory foresee being revised if needed for implementation. But in practice there are often significant barriers to doing so, e.g. achieving consensus on adjustments, time constraints and procedural requirements. This is particularly difficult for programme theories that were developed in a participatory manner, and there even seems to be an inherent paradox: the more people participate in their development, the more difficult it is to revise this consensus later on during implementation and to adapt an intervention to changes (Bakewell & Garbutt, 2005). Therefore programme theories are often not up-dated, but tend to be fixed and thus prevent learning and adaptation.

Adaptive management can be supported if the review of programme theories duly pays attention to the following:

- Facilitate the adaptation of programme theories: changes are normally foreseen during a formal review, or in response to important changes in circumstances. But the procedures might be quite cumbersome and time-consuming and prevent revisions from being carried out in due time. Adaptation might also have negative connotations in an administrative culture, which is averse to change – or even learning.

- Orchestrate and approve changes: changes can have an effect down-wards as well as up-wards in the causal chain, so it is advisable to consult upwards and downwards with affected partners about proposed changes, in order to address their full implications. All changes should be endorsed by the same authority that has approved the former version. Different authorities might have to be involved depending on the concerned level of effects, e.g. changes at purpose or goal level might require approval by a higher authority.

- Pay attention to emergent phenomena: acknowledging that not everything can be anticipated is an important ingredient for staying attuned to reality. The minds of partners and decision-makers should be sufficiently open and attentive to be able to identify the appearance of new factors or observe relevant changes, which are often incremental and thus easily overlooked (at least when they first emerge).

- Deliberately look for surprises: in order to avoid the lock-in or 'tunnel effect', interventions should try to look beyond their intended routes. During monitoring and periodic reviews, specific attention should be paid to capturing unexpected and unplanned effects.

When reviewing progress made towards outputs and outcomes, differences from original plans should not *a priori* be seen as negative 'deviations' nor treated in an isolated manner, but should be linked to the programme theory. These differences should be valued and explored in more detail, particularly taking into account that:

- Analysing differences in output (as well as outcomes) can help to assess the appropriateness of an intervention in view of the given operating context (e.g. framework conditions, needs of target groups, interests of implementing partners).

- Exceptions, discontinuities, unforeseen results and side effects are valuable sources of information. They can provide useful clues for relevant internal/external changes, new challenges,

innovative or 'informal' ways of handling situations, all of which could help to improve implementation.

Developmental evaluation (Patton, 2010; and the chapter in this volume) focuses in particular on evaluation designed to inform adaptive management. Programme theory for evaluations needs to be either highly adaptable and reviewed often, or be at a higher level of abstraction that remains constant despite changes in the details.

Conclusion

Programme theory evaluations of equity-focused interventions need to pay attention to how the interventions are supposed to work, whether or not they have differential effects, as well as whether or not they have achieved desirable impacts. Using programme theory thoughtfully in this way will increase the likelihood that evaluations of individual interventions will gather valid and useful evidence – and increase its utility for learning from these interventions for future work.

References

Ambrose, K. & Roduner, D. (2009). *A conceptual fusion of the Logical Framework Approach and Outcome Mapping*, OM Ideas Paper No. 1

Bakewell, O. & Garbutt, A. (2005). *The use and abuse of the Logical Framework Approach. A Review of International Development NGOs' Experiences.* Sida.

Bamberger, M, & Segone M. (2011). *How to design and manage equity-focused evaluations.* New York; UNICEF.

Church, C. & Rogers, M.M. (2006). *Designing for Results: Integrating Monitoring and Evaluation.* In: *Conflict Transformation Programmes.* Washington, DC: Search for Common Ground.

Davies, R. (2005). *Moving from Logical to Network Frameworks: A modular matrix approach to representing and evaluating complex programmes*.

Earl, S., Carden, F., Smutylo, T. (2001). *Outcome Mapping: Building Learning and Reflection into Development Programmes*, Ottawa: IDRC.

Fernandez, J.; Gandarillas, E.; Polar, V.; Fuentes, W.; Almaza, J. & Quiroz, C.A.; (2006). Seguimiento y Evaluación Participativa en organizaciones locales. Una herramienta de control social. Cochabamba, BO. Fomentando Cambios.

Forss, K., Marra, M. & Schwartz, R. (2011). *Evaluating the Complex: Attribution, Contribution, and Beyond.* Piscataway, NJ: Transaction Press

Funnell, S. & Rogers, P. (2011) *Purposeful Programme Theory: Effective Use of Theories and Change and Logic Models.* San Francisco: Jossey-Bass/Wiley.

Glouberman, S., & Zimmerman, B. (2002). *Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?* Ottawa: Commission on the Future of Health Care in Canada, Discussion Paper 8, 2002.

Green D. (2011). 'How can theories of change help in working with the private sector?' From *Poverty to Power* blog. http://www.oxfamblogs.org/fp2p/?s=archetypes&x=0&y=0

Hambly Odame H. (2001). *Engendering the Logical Framework,* International Service for National Agricultural Research (ISNAR)

Hummelbrunner, R. (2010). 'Beyond logframe: Critique, Variations and Alternatives', in *Beyond Logframe; Using Systems Concepts in Evaluation*, Tokyo, FASID.

Kurtz, C. F., & Snowden, D. F. (2003). 'The New Dynamics of Strategy: Sense-making in a Complex and Complicated World'. *IBM Systems Journal*, 2003, 42(3): 462–83.

Leeuw, F. & Vaessen, J. (2009.) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*, World Bank, Washington, D.C.

Mertens, D. (2008). *Transformative Evaluation*. New York: Guildford Press.

Patton, M. Q. (2010). *Developmental Evaluation*. New York: Guildford Press..

Pawson, R. (2006). *Evidence Based Policy: A Realist Perspective.* Thousand Oaks, Calif.: Sage.

Pawson, R. (2008). 'Invisible mechanisms'. *Evaluation Journal of Australasia*, 8(2), 3-13.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Thousand Oaks, Calif.: Sage..

Paz R., Dorward A., Douthwaite B. (2006). Guia metodológica para la evaluación de impacto pro-pobre de proyectos de innovación tecnológica en agricultura de pequeña escala. Centro para el Desarrollo y la Reducción de la Pobreza. Centro de Política Medioambiental. Imperial College London.

Practical Concepts. (1979). *The Logical Framework: A Manager's Guide to a Scientific Approach to Design and Evaluation*. New York: Practical Concepts.

Ramalingam, B., Jones, H., Young, J. & Reba, T. (2008). *Exploring the Science of Complexity: Ideas and Implications for Development and Humanitarian Efforts*. ODI Working Paper 285. London: Overseas Development Institute.

Rogers, P. J. (2008). 'Using Programme Theory for Complicated and Complex Programmes.' *Evaluation: the international journal of theory, research and practice*, 2008, 14 (1), 29–48, 2008.

Weshthorp, G. (2009). *Understanding negative impacts of early years intervention programmes: realist methodologies*. Paper presented at the conference of the Australasian Evaluation Society, Canberra. http://www.aes.asn.au/conferences/2009/documents/Gill%20Westhorp.pdf

White, H. (2009). *Theory-based impact evaluation: Principles and practice*, Working Paper 3, International Initiative for Impact Evaluation, New Delhi

Williams, B. & Hummelbrunner, R. (2011). *Systems Concept in Action – A Practitioner`s Toolkit,* Stanford University Press

World Bank. (2002). *Beneficiary Assessment.* http://siteresources.worldbank.org/INTPSIA/ Resources/490023-1121114603600/beneficiary_assessment.pdf

CASE STUDY AND EQUITY IN EVALUATION

*Saville Kushner, Professor of Public Evaluation,
University of the West of England, UK*

Introduction

It is not the job of evaluation to change the world. Indeed, the independent evaluator should be constructively sceptical about the claims of those who do seek to change the world. The evaluator's job is to identify and report on programme quality, to help people come to terms with the complexity of programmes, and to feed the judgements of those who have decisions to make about programmes. To report on programme productivity (impact, results) is often an important responsibility of the evaluator, but to focus on that to the exclusion of programme quality is to confine evaluation to a narrow range on the spectrum of possible roles.

This is not to say that evaluation should be detached from the programme it observes. At best, evaluation supports the moral aims of a programme which seeks to enhance democracy and equity. An evaluation should add to the quality and energy of programmes. This is especially so in relation to development programmes where issues in good governance almost always provide a context to the programme's ambitions. Implementing the Cold Chain or disaster relief or a maternal health programme is more effective by degrees the more democratic a client society is. Evaluation can and should model democratic process through its own conduct – be the conscience of society in its deliberations over social planning for new futures. Indeed, when Cronbach (1981, p.2) came to write his seminal '95 Theses' for evaluation, the first was:

'Evaluation is a process by which society learns about itself.'

The significance of this volume lies in its advocacy of evaluative enquiry, not just to evaluate programmes of equity, but to stand for equity itself, through its own interactions – equity-*based* as well as equity-*focused* evaluation. Of course, this places limitations on evaluation impartiality, makes it something of an ideological process – especially when it is conducted in societies which resist democratic development and equity-based solutions to social issues. Evaluation that

promotes equity may find itself taking sides – say, with a citizenry and against an economic or political elite who might stand to gain from inequity. We need look no further than the United Kingdom where we see a gradual shift from a progressive to a regressive taxation system and with tax-breaks for wealthy businesses matched with cuts in social welfare – evidenced by a consistent erosion in the country's Gini Coefficient since the 1970s. In such a context, evaluation, free to examine issues of equity and rights, finds itself in increasing struggle to retain its independence (Elliott & Kushner, 2007).

This chapter contributes to the volume with an approach to evaluation that makes it an equitable process. Its focus is methodological, and the methodological solution to the problem of achieving equity is case study. This was designed in the decade of the 1970s as a methodological expression of democratic intent, not as a technical solution to generating qualitative data (see Simons, 1980, Elliott & Kushner, 2007). We look for equity in evaluation in the following places:

- In its purposes: evaluation should serve multiple logics, there is rarely a singular programme logic. All programmes have multiple ('broad') aims (Weiss & Rein, 1969) and equity demands that all aims are honoured. Indeed, it is often the dynamic relation between competing purposes that characterises a programme and its qualities (e.g. where a ministry, its constituent institutions and those institutions' practitioners each have different views and priorities in respect of an intervention).

- In its design: evaluation design should include the questions and objectives of all stakeholders – especially the citizen and the programme worker. To assume a programme has only one set of objective is to violate a principle of equity. There is no good reason (though there are many pragmatic reasons) to elevate official programme objectives over others.

- In its access: access to evaluation should offer no privileges and should be available to all. This requires of the evaluator that they have an open field presence and that they offer no private briefings.

- In its conduct: evaluation should treat all stakeholders in the same way – a minister, a manager, a teacher, a pupil, a parent – equity demands that they all share the same level of rights in evaluation, and the evaluator should enshrine that in his or her conduct. This stems from the principle enunciated in the early

moments of programme evaluation, that *'people own the data over their own lives'*.

- In its negotiations: again, equity demands that there are no privileges in evaluation. If there is an expectation that an evaluation report is to be negotiated, then all constituent groups in a programme – indeed, all those represented in the evaluation report – have the right to negotiate over the content and conclusions of the evaluation. This also stems from the individual ownership of data.

- In its reporting and narrative strategies: for evaluation to be accessible to all implies that it is reported in terms that allow all to interact with its reports in meaningful ways. This often means reporting in narrative form, on an equity basis.

- In its utilisation: Once more, the principle of equitable access to evaluation implies the universal right to be able to use evaluation in a constructive way as an aid to self-determination.

These are hard messages for evaluation – especially for sponsors of evaluation. It is often the case that an evaluation is sponsored according to the urgencies and priorities of the sponsor – most frequently the administrative system. However, this, again, violates the principle of equity. In an early and seminal paper, MacDonald (1987) argued that evaluation should be thought of as *"sponsored but not bought"* – i.e. enjoying the capacity to sponsor an evaluation should confer no special privileges. However, it represents the scale of the challenge for achieving equity in evaluation, and for evaluation to stand as the conscience of society (House 1973).

Evaluation, equity and the power of narratives

> *"There is an unbridgeable, but largely unappreciated gap between the neat rationality of development agencies' representations which imagine the world as ordered or manageable and the actualities of situated social practices…The result is that the overlap of developers' and local discourses does not lead to improved communication, but to strain on those locals who are involved in both…"*
> Hobart, 1993, p. 16

A key source of social inequity is the asymmetry between what we can call 'the official narrative' and what we might call 'experiential narratives'. Official narratives are explanations of social issues as expressed by government and its agents. Experiential narratives

are expressions of priority and need as expressed by civil society and those whose work is directed at its improvement – practitioners – teachers, social workers, police officers, clinicians. Government has resources to assert its preferred narratives over those of civil society, and this is what frequently confronts evaluators who seek to explain the context of interventions. This is a simple power asymmetry and a structural inequity, illustrated in Hobart's quote above. The asymmetry arises out of differential capacities to act in meaningful and effective ways. They are better described as 'asymmetries of agency'.

How do evaluators negotiate that space between *'neat rationality'* and the messy, '*actualities of situated social practices*' – i.e. how does evaluation strive for equity in such an inequitable situation? How do we improve mutual understanding and communication between government interventionists and civil society – make the former more responsive (accountable to?) the latter? It is a common observation in advanced industrial societies that there is a tension between policy and practice – arising from a mutual lack of understanding, competing values and interests, different ways of seeing, differing accountabilities, and so on. A great deal of evaluation effort has been spent on understanding and mediating these tensions – using evaluation as an instrument for discovering more equity in the distribution of knowledge – equity based on mutual understanding. This use of evaluation is little practiced in the development field. Prominent among methods for creating equity through mutual understanding has been through evaluation case study, and this is what this chapter is about. The simplest definition of case study I will explore is, *'the study of context and contingency as a determinant of action'*. (Contingency refers to how things are interrelated in dynamic ways).

Without communication and mutual understanding there is restricted space for negotiation of realities and priorities and so limitations on equitable access to knowledge of development interventions. At best, case study provides a social and political space for information exchange, which allows for building social consensus – bridging that gap between international and political elites and communities and making their access to knowledge more symmetrical. Evaluation can promote equity where it equalises access to the criteria over what counts as public value.

We have to know how we see a programme of intervention

The question of how evaluation draws from the full spectrum of possible roles offers something of a response to critiques of international development evaluation such as that of Bollen et al (2005), who argued that among the 240 evaluation reports they reviewed there was widespread evidence of lack of methodological sophistication (especially validity), absence of an awareness of the contexts in which evaluation took place, and reliance on limited and often flimsy data sources. Another important critique comes from Marsden (2003), Research Director at INTRAC, who argued for 'ways of seeing' (an interesting definition of evaluation) that are 'cultural' and capable of capturing and nurturing flexible, alternative, sometimes network forms of social organisation, and that depart from 'mechanistic' ways of seeing poorer societies and their solutions. His argument is especially demanding in international development where we routinely apply methodologies derived from North/Western, Judaeo-Christian political culture to the diverse realities of indigenous and other peoples, inevitably widening the gap between 'neat rationalities' and the messiness of lived experience – that fundamental source of social inequity. Picciotto (2007) argued that evaluation in the field of international development is too narrowly conceived and should embrace wider political contexts than those specified in a brief. Such critiques mirror vigorous debates about evaluation that have been held in countries of the North for the past 30 years and which have helped mature this democratic practice. There is no reason to deny the developing world such deliberations. So let us extend evaluation to that part of the spectrum populated by case-based, narrative approaches to building equity into evaluation – i.e. equity derived from equal access to evaluation. First, we need to look at an evaluation in its context.

Figure 1: Programmes and contexts
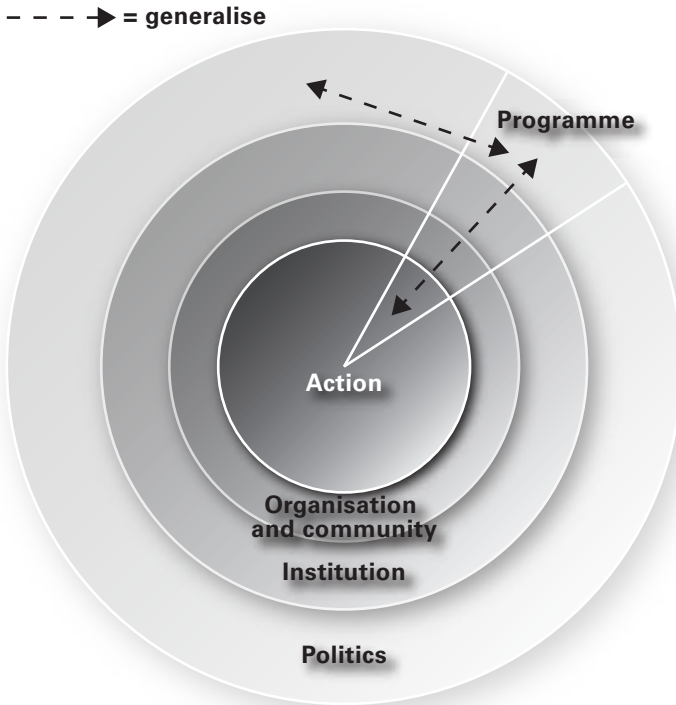
Source: Kushner, S. & Adelman, C. (2006)

Figure 1 is a simple, two-dimensional view of a sector – let us say, health care – and given the limitations of two dimensions, it defines a typical programme structure. Each of the circles represents a different context with its own practices, values, accountabilities, etc. Each boundary between levels involves negotiation and the interplay – sometimes a clash – of values and interests: organisational managers with practitioners; policy shapers with donors; service users with managers – each has to understand the other – often with little available information. The two circles at the centre of the diagram coincide with civil society and its organisations (CSO) – and with lived realities. Each level is in dynamic relationship with others – together, they all determine action. We can imagine quite easily that power frequently accumulates at the outer edges of the diagram, leading to a key source of inequity – i.e. asymmetries of agency. How does evaluation confront that?

This structure can be mapped onto the design of an evaluation case study. The 'cake-slice' represents a single programme of intervention – in our example, let us say, a *child nutrition intervention*. It has a context for action – professional practices among health visitors, nutritionists and clinical workers. These are located in specific organisations – primary health care clinics, schools – which are, themselves, located in institutional frameworks – professional and governance arrangements including municipal government and professional associations. Finally, all is encompassed by a political system, topped by ministries of health and finance, but also including international agencies and others who advocate health and fiscal policy. The rest of the 'cake' represents all other parts of the health sector, and this will include other health programmes of intervention.

The first demand that such a view of a programme makes on the evaluator is to move across these levels and boundaries to construct a sampling frame. Why? The purpose of any programme is to unite efforts around agreed goals. However, given the reality that these are most likely fragmented systems with disagreements between some fragments, consensus – based on equitable access to evaluation information – is an *accomplishment*, not an *input.* Evaluation provides political and ethical space within which information exchange creates equity and stimulates democratic argument. Each level, then, offers distinct opportunities for data generation to the evaluator, but also places a demand on the evaluator to help each constituency to understand the priorities and purposes of the others. The evaluator has to travel 'up' and 'down' the cake-slice in order to foster public conversation about what is worthwhile as revealed by the intervention, and to ensure distributed access.

Practitioners and managers are mostly confined to their level – condemned to their perspective – and this is true, but often less so, for international agencies including the United Nations. From this comes the first obligation of the evaluator, to describe the *whole* programme in ways no other observer can – to reveal it. It is for this reason that some writers on evaluation case study characterise the methodology as *the study of contingency* – i.e. analysis of how programme elements interrelate, often in dynamic ways – for example, how action depends on context, how institutional development depends on professional development, how practice arises out of or is constrained by values and interests – but principally, how one level depends upon another to realise its ambitions. This demands direct observation (portrayal) of the programme and sampling its interactions (e.g. meetings, field relationships and conversations, workshops, documents,

reflections). Having completed the analysis of this programme, the evaluator can generalise to other parts of the sector – to look, say, at public health, acute care, maternal health programmes. That is the arrow leading from the 'cake-slice' to the rest of the cake. (Take a moment to look at that).

Case study evaluation, in this sense, plays to the demands of sector-based approaches to poverty alleviation and rights promotion. The shift of attention among multilateral development organisations from field-based to policy-orientated interventions risks losing the connection between the two – losing situational understanding of how policy, organisation, professional practice and community priorities are all in dynamic relation with each other – and how these tensions are resolved in practice. Case study can remake that connection. This provides the firm footing for dialogue over rights and how results do and do not contribute to their realisation.

We could also say, however, that the 'cake-slice' is itself the health sector and the rest of the cake is the remainder of all public sectors – education, criminal justice, urban planning, etc. Now, the evaluator can generalise from the observation of the health programme to other social programmes. For example, having understood how policy and practice are related in the context of a health innovation the evaluator is equipped with certain insights in innovation and change that may be transportable to other contexts such as community policing, child protection and education where similar relationships exist. In analytical terms we would say that the evaluator has developed a *theory of contingency* which can be applied to other programmes.

Case study gives us a more systemic and dynamic view of policy and public value

We should not imagine that the outer circle – 'Politics' – is where we would locate Policy. We need to think of policy as a property of the whole system, created out of its interrelationships and shifting and changing with the dynamic relation of system parts and levels. In the real world, policy is *shaped*, not *made*. For example, national health policy may insist on certain birthing practices; but over time, mothers' and midwives' cultural preferences will exert influence on policy. Part of policy-shaping involves a probable clash of values between, for example, a political elite, a professional group and citizens – in our case, mothers. The important point, however, is

that we see programmes, not as concrete resolutions of ministerial decisions, but as *sites* within which we can observe policy being shaped and value being competed for. It is the reality that policy is shaped that allows us to deploy evaluation as an instrument for equity, since the iterative shaping process allows for citizen agency. This demands of case study that it reveals the policy-shaping process as it:

- observes programme interactions directly – e.g. interactions between programme managers and its practitioners, interactions between health clinicians and mothers;

- is reported in a narrative style that makes it accessible to all;

- systematically compares the programme's intentions with its actual performance, and compares the policy behind a programme with the values of programme practitioners and citizens;

- documents the different intentions and interpretations of the programme held by diverse stakeholders;

- documents political, cultural, organisational and community contexts as well as programme action;

- shares information across stakeholder groups so that proper argumentation and deliberation can be had over what is worthwhile and how to place value on the intervention;

- documents programme interactions in such a way that reveals how individuals translate programme aims into action.

Describing, analysing and understanding the implications of policy-*shaping* is a key task for the case study evaluator who needs to understand the sources and consequences of unequal access to information, and power asymmetries in setting the criteria against which interventions will be judged. Observe a teacher, a health visitor, a child protection commissioner at work and you are observing the shaping of social policy, not just its implementation. There is little such direct observation of the *process* of producing results in development evaluation. Indeed, the practitioner is the Cinderella of the field.

Data sources in case study are diverse

What data does the evaluator collect in her journey 'up' and 'down' these political systems? Here is a typical check-list, keeping in mind our example of a child nutrition intervention.

(A)   First order data

The political level

- Analysis of government policies and international agreements (as appropriate) regarding child nutrition, poverty reduction, agricultural development and extension, cash-transfer schemes, etc. These are analysed to identify coherence and continuity.

- Interviews with key respondents in policy-shaping processes at national and other levels exploring political values and priorities, but also personal values of key players. What social values underpin models of nutrition and wellbeing?

- Comparative analysis of nutrition-related budgets with nutrition-related policy – history and present – to identify continuity or discontinuity.

The institutional level

- Mapping of child development-facing infrastructure including a children's workforce.

- Assessment of levels of professionalization of a children's workforce.

- Quality and reach of professional training systems including engagement of tertiary education sectors – how sensitive are training curricula to local and cultural practices?

- Conceptual quality of innovation and change theory underpinning infrastructure development and institutionalisation.

- Interviews with key respondents with responsibilities for infrastructure development.

Organisational and community level

- Case studies of organisations through which the intervention operates – to capture philosophy, constituency, range of practices, levels of professionalization, levels of interaction/linkage/integration.

- Critical review of methodologies for measuring/treating malnutrition – e.g. what reliance on measures of stunting/wasting and to what distorting effect?

- Review of content, focus and quality of training.

Action level

- Interviews with a sample of practitioners delivering the intervention to identify values-base, priorities, evaluations of the intervention, predictions, commitments to practice.

- A sample of observations of interactions in the intervention between practitioners, families and children including knowledge/power relations, language, technique.

- A sample of ethnographic interviews with families and community leaders focused on life-styles, agricultural practices, patterns of consumption and nutrition practices.

(B) Second-order data

- Comparative analysis between intervention aims and objectives, political values and practitioner values.

- Comparative analysis between the development of policies and corresponding developments in social budgeting – and mapping of that onto organisational resourcing.

- A further programme of interviews with practitioners and their managers based on observations of the programme's interactions (exploring rhetoric/reality gaps).

- Mapping of aims and objectives of participants/stakeholders in the programme and comparative analysis of these with official aims of the intervention – whose objectives drive/ constrain action.

- Comparative analysis of change theory at different levels of the system.

This indicative list constitutes an evidence base on which to understand the following:

- how people at all levels 'fit into' the intervention – but also, how the intervention finds functional fit in the lives of its stakeholders;

- potential for differences and fragmentation of effort set against the potential for forging consensus over values, priorities and practices;

- assessment of the coherence of the initiative and its theory(ies) of change in the contexts in which it is implemented, along with a critical review of alternative views in the cultural context of malnutrition and its treatment;

- Analysis of variation of practice and outcome across intervention sites.

Case study helps to link evaluation and governance

"In practice, evaluation can be…democratised by extending evaluative choices to all groups and by extending pubic evaluation to all public choices."
House, 1980, p.101

The proximity to both action and policy contexts demanded by case study is ideal for generating first- and then second-order data as we see above (i.e. through return visits), checking data against observed realities and negotiating improved understanding with programme participants – valuing the judgement of programme participants. In fact, observation-based interviewing emerged in this context – conducting direct observation of programme activities and then inviting the actors to reflect on them critically. Participants (practitioners, citizens, young people) are not invited to participate only on the basis of equity and fairness, but because their personal experience and judgement are vital to understanding the programme and its potential in ways not accessible to the evaluator (especially with regard to attribution). In a school curriculum project or a community policing initiative it is the teacher and police officer who have the expert view derived from the daily challenge of making sense of programme goals, policy aims and community realities and priorities – and, as we have seen, this is an important element in the policy-shaping process. Indeed, it is the practitioner voice which is the most significant loss in development evaluation. This is collaboration, not participation.

More important, still, is that the case study allows the evaluator to shift from asking *retro*spective questions ('why did you do this?' 'how did this come about?') to asking *pro*spective questions ('what would make a difference?', 'how should this be better managed?'). They are the prospective questions that allow debate about how to get public value to emerge.

Prospective questioning, making transparent and publicly accountable the question of how to value a 'result' treats programme participants as a decision-making resource, but also makes evaluation an equitable activity in itself – i.e. conferring information and access rights and honouring the judgement of civil society as well as official agency. Giving programme workers and citizens the opportunity to

reflect on what *might be*, challenges the concentration of power in programmes by shifting control over the criteria against which the quality and success of the programme is to be measured. This is how case study evaluation grounds the definition of, and claiming of rights, and subjects to scrutiny the programme's view of what counts as public value. A programme defined by a single set of (official) goals denies alternative definitions of what the programme is and might be, and so restricts deliberation over rights and value. It also uses judgement criteria from the same sources of official values and interests that first gave rise to the programme, denying other stakeholders their own right to make what might be quite different judgements about the programme[1]. Case study creates an ethical site within which the evaluator's obligations to programme stakeholders are made transparent and subject to sanction (e.g. people can refuse to talk to evaluators and should have the means to challenge an evaluation finding). In such a site people will naturally assert their right to make their own judgements of the programme, and the evaluator is required to respond on the grounds of equity. The fundamental ethic of evaluation conducted at close proximity is based on the evaluator's accountability to stakeholder constituencies.

It is in this way that case-based programme evaluation can be an instrument for enhanced public accountability and equitable access to knowledge, and an opportunity to involve civil society and its organisations in debates about policies designed to shape their lives and work – *engagement*, not *participation* – based on prospective rather than retrospective questioning. '*How and why do we value what this intervention offers?*' is a key evaluative question. This is an opportunity to invert accountability relationships: as well as holding practitioners to account for their success at realising the practical aims of policy, programme managers and political elites could be held to account for being responsive to the complex challenges of public service.

The final link is made between programme evaluation and governance, allowing for the pursuit of equity – at least in the confines of the evaluation project. House & Howe (ibid.) argued that this link is vital to evaluation practice:

---

1   A contemporary example of this is the evaluation milieu of the enormous (at one stage more than $1bn per year) federal US programme, *Drugs Abuse Resistance Education*. Evaluations using 'official' methods and criteria – such as that of the Surgeon General of the US – declared the programme was failing in its objectives. The programme itself and many of its stakeholders (including the police) have insisted that, against their practical criteria the programme is effective in improving their lives.

"...background conditions for evaluation should be explicitly democratic so that evaluation is tied to the larger society by democratic principles argued, debated and accepted by the evaluation community. Evaluation is too important to society to be purchased by the highest bidder or appropriated by the most powerful interest." [p.4]

Case study, Results-based management and agency

Under Results-based management, the principal criterion for validating a development intervention is that its results can be pre-specified, i.e. that the objectives are accurate predictors of outcomes. There is a reduced need for the intervention to show good correspondence with context, such as the mechanisms of change and causality or, indeed, to show ethical integrity – contextualisation might even be thought to *undermine* validity (Gerring, 2007). This is sometimes an unwelcome impact of the international commitment to Results-based management (RBM) which insists on the pre-specification of results (commonly, as far as five years ahead in the case of multilateral programmes of cooperation). As currently practised, RBM *at national and aggregate levels* diverts attention from causal processes and refocuses resources from *learning* (adjusting to context) to *justification* (for accountability). It has a tendency to concentrate control in donor countries. The principal focus of concern is not programme quality, but programme productivity which stands as a surrogate for quality. RBM – and I emphasise, *as commonly practised* – overwhelms contextual considerations (see, Ortiz et al, 2004 for an example).

This risk of missing the complexities and uncertainties of causality leads to a departure from the fundamental task of any evaluation, which is to describe programme process, foster exchange across programme constituencies and to feed into judgements of programme quality. It is not helpful where an evaluation talks to the reader *about* a development programme without allowing the reader to *see it* directly – too few evaluations of development interventions include observational data giving the reader direct access to the programme. One of the effects of this is to diminish the capacity (the right) of programme participants, citizens and donors – but also evaluation sponsors and political elites – to make their own judgement of a programme's merits. What is lost with an exclusive reliance on results is a public debate about what counts for quality in programmes of intervention, and what counts as a valid and worthwhile result – i.e. what has public value.

There is yet another effect of a loss of programme realities. The lack of descriptive, case-based data means that through our measurements we know how mothers and children, say, fit into development interventions; but we know little of how interventions fit into their lives. It is to this scenario that we now turn.

It is curious and worrying to observe that the overwhelming focus of development interventions and their evaluations is on the poor. There is little, if any, documenting of the lives of the wealthy in developing countries, and few studies of elites. If the world to international multilateral agencies appears 'ordered and manageable' (Hobart, ibid.), this is to some extent, due to the partial view of society created by programmes and their evaluations – with a great deal of complexity and cultural diversity left out. That principle of case study – that we seek to embrace all variables for the purposes of comprehensive understanding and not to control for variables – is violated to the point where proper understanding of programme interventions is compromised. Nor do we have a robust foundation on which to practice equity. Not only do we not know how interventions fit in the lives of the poor, there is typically little knowledge available of how interventions fit into society as a whole. One of the unrealised promises of development evaluation is its capacity to engage rich and poor classes in dialogue about public value, about social responsibility, about the balance between personal and collective wealth – about equity. Two examples will suffice:

(a) The irrationality of wealthy, gated communities living amid squalor and insecurity demands a resolution. Debates about what counts as 'wealth' are long overdue – for example, the balance between private accumulations of artefacts (cars, houses, jewellery) and public artefacts (good roads, lighting, public amenities).

(b) Comparisons of the effects of wealth/poverty on mothering and child health – what can wealthy mothers learn about 'mother's medicine' from poor mothers – and *vice versa*?

The most serious information losses are those related to an understanding of causality (how and why results are and, at least as often, are not generated), and to public accountability (i.e. the lives and contingencies of 'duty-bearers' as well as those of 'rights-holders').

There are particular instances of this absence of contextual analysis which can be seen in RBM:

1) *Ends-means rationality*, in which the ethical justification of method and process is given by the 'goodness' of the outcome – in the international development field, usually achieving Millennium Development Goals. There may be harm or controversy during the intervention to meet these goals (e.g. occasional infant deaths resulting from mass vaccination programmes), but this is overwhelmed by the universal good that is eventually served, and so is of secondary interest. Such utilitarian thinking (See House, 1993) continues to be promoted by political elites in donor countries.

2) *Absence of a theory of contingency*. This rests on a belief that cause and effect in an intervention are sequential and connected *logically*, not *empirically* – for example, the political assumption that things will happen because they are planned to happen and because resources map out a pathway for objective to move smoothly to result. Hence, RBM is practiced with the use of 'Log Frames' and 'Results Matrices' which specify collaborative effort among partners, for example, but tend to avoid intervening politics in those partnerships. One frequent problem is that results matrices rarely address the frequent real-life tension between political values and professional practitioner values.

3) *The separation of the intervention from time and locality* – mostly through the long lead-time pre-specification of results. 'Technicism' and RBM deal with the general case – indeed, an important element of validity is the suppression of local timeframes and the fact that the intervention follows the chronology of its own internal accountabilities and is not compromised by the realities of the case. MDGs, for example, are to be met by all regions and countries by 2015 and international organisations publish their own milestone-deadlines along the way, irrespective of the pace of social change in target countries. Hence, variables (interactions with context) are to be 'controlled for', 'neutralised', or 'isolated' so as not to impede results achievement – whereas they are precisely those extraneous contextual variables which give an insight into how MDGs, in this case, work with or against the grain of local factors. A technology – a washing machine and a global MDG indicator, for example – is expected to function the same and to hold the same meaning no matter what the context.

4) *An assumption that the quality of an intervention is given by its productivity*: under RBM a programme that achieves its predicted results (and eventual impact) is assumed to have quality. However, results are surrogate measures of quality and not direct measures. There are too many intervening variables between programme action and programme impact.

Each of these resistors is challenged by case study defined as a study *in* context, and the study *of* context. In contrast, then:

(a) Case study evaluation assumes means-ends rationality, in which end-results are subservient to the realities faced by the intervention, and its possible outcomes are at least partially determined by conditions and priorities within the case. MDGs, that is to say, would have to be reinterpreted and accommodated at the level of community, whose preferences should drive the intervention. All interventions should start with case-based evaluation.

(b) Case study *is* the study of contingency (Stake, 1967, 2004), in that it assumes the overriding goal of understanding complex interactions that intervene between presumed causes and observable effects – i.e. not *what* programmes achieve, but *how they work*. MDGs assume a different kind of significance in case study evaluation: rather than goals and stipulations they become *sites* within which we negotiate means and priorities. An acceptance of contingency allows us to think more realistically of policy in terms less of decision-making by a political elite, and more of policy-shaping communities.

(c) for the purpose of integrity, case study seeks to embrace the broadest range of variables possible so as to enhance the quality of its insights and to improve its explanatory potential – i.e. to be valid, case study has to explain the relationship between context and intervention.

The potential for case-study approaches to development evaluation is high, the urgency equally so. Insufficient work has been done to ground MDGs in the priorities and meanings of communities at whom they are targeted. The challenge of democratically evaluating global programmes such as the MDGs and RBM is insufficiently explored. *"To commit to the MDGs implies that the realism and fairness of the goals themselves and the soundness of its underlying theories of change constitute legitimate objects for development evaluation,"* wrote Picciotto [p. 512]. 'Realism' and 'fairness' are, as

we have seen, the stuff of case study, for they are only researchable in context – since they are defined by context. The reach implied from fairness-testing (i.e. with communities), to feasibility (with delivery institutions), to testing the theories of change (development agencies and governments), again calls for case study approaches, i.e. comprehensive, multi-level, judgement-based and system-analytical methodologies. Franklin (2007) in his critique of RBM, emphasises the grounding of evaluative enquiry in civic, deliberative process – again, reflecting evaluation theorists confidence in case study, and focusing on the 'fairness' agenda – here he defines a Democratic Evaluation site:

> *"Communities need safe spaces in which they can discuss sensitive topics, where all people feel confident enough to exchange opinions and listen to each other, and where people with different and sometimes conflicting points of view or interests can agree on what needs to be done to change things they feel need to be changed. International development agencies are not usually equipped to create such spaces."* [p.422]

De Lange (2003) argues that in a world attempting to make North-South relationships more 'symmetrical' (partly, some argue, through the Paris Declaration) RBM can have a tendency to reinforce asymmetries – that, if used, RBM should be subsumed under participatory, client-centred partnerships where Southern values and preferences are not subverted:

> *"[donors] use results…as feedback to adjust their policies and programmes. They then discuss or negotiate these changes with their Southern partners or impose them. If RBM-based evaluations are conducted in such a setting they are very likely to enhance existing inequalities in the relationship and prevent Southern partners from learning."*

It is feasible, and desirable, for RBM to serve democratic and rights-based aims in the terms implied by De Lange. But this would involve placing RBM *within* a local framework of public value and not prior to it or extrinsic to it; it would involve local public deliberation over the meaning and the significance of results against the particularities of context, and it would involve continual review of the merits or results as evaluation data comes in.

In a context of global accountability and global goals the democratic legitimacy of international agencies is sometimes stretched thinly enough to make it problematic to maintain currency at the level

of community and municipality. In one sense, MDGs represent a global asymmetry (not that they are not worthy in themselves – this is not the argument). The sprawling, loosely-coupled nature of multilateral agencies like the UN – the overseer of MDGs – intensifies the potential for systems fragmentation, stretching even more the distance between the citizen and the political elite, thereby intensifying the urgency to bridge the distance, to strengthen public accountability based on deliberative processes, and to inform those international agencies of how their rights advocacies fit into the contexts in which they intervene. The potential, as we have seen, is for case study approaches to integrate local, national and global perspectives and to ground discussions over rights – as well as providing space for civil society to claim their rights and assert their priorities. Through the use of case-based evaluation, international agencies can reinforce the legitimacy that allows them to advocate for democratic governance.

References

Bollen, K, Paxton, P. & Morishima, R. (2005). *Assessing international evaluations: an example from USAID's Democracy and Governance programme* in American Journal of Evaluation, 26(2), pp. 189-203.

Elliott, J. and Kushner, S. (2007). 'The need for a manifesto for educational programme evaluation'. *Cambridge Journal of Education*, 37 (3). pp. 321-336.

Franklin, T. (2008). *Reaching the Millennium Development Goals: equality and justice as well as results* in Development in Practice, 18:3, pp. 420-423.

Gerring, J. (2007). *Case Study Research: Principles and Practices*, Cambridge: CUP.

Hobart, M. (1993). *Introduction: the growth of ignorance* in Hobart, M. (Ed.) *An Anthropological Critique of Development*, London: Routledge

House, E. (1973) 'The conscience of evaluation' in House, E.R. (Ed.) *School Evaluation: the Politics and the Process*, Berkely, Cal: McCutchan

House, E. (1980). *Evaluating with Validity* Thousand Oaks: Sage.

House, E. (1993). *Professional Evaluation* Thousand Oaks: Sage.

House, E. & Howe, K. (2000). *Deliberative democratic evaluation* in K.E. Ryan & L. DeStefano (Eds.) *Evaluation as a Democratic Process: Promoting Inclusion, Dialogue and Deliberation*, New Directions for Evaluation (AEA) No. 85, SF: Jossey-Bass, pp. 3-12.

Kushner, S. & Adelman, C. (2006). *Programme Evaluation: a Democratic Practice* for the AERA Handbook, *Complementary Methods for Research in Education,* NY: Laurence Erlbaum.

Lange, P. de (2003). *Dilemmas in RBM-oriented evaluation practices: back to the basics?'* keynote presentation to the 5th *INTRAC International Evaluation Conference*, Netherlands.

MacDonald, B. (1987). 'Evaluation and the Control of Education' in Torrance, H. & Murphy, R. *Issues and Methods in Evaluation*, London: Paul Chapman, pp. 36-48

Marsden, D. (2003). *Rights, culture and contested modernities*, keynote presentation to the *5th INTRAC International Evaluation Conference*, Netherlands.

Ortiz, E.F., Kuyama, S., Munch, W. & Tang, G. (2004). *Implementation of Results Based Management in United Nations Organisations*, Joint Inspection Unit, Geneva: United Nations – available in pdf: http://www.unjiu.org/data/reports/2004/en2004_6.pdf

Picciotto, R. (2007). *The new environment for development evaluation* in American Journal of Evaluation, 28:4, pp. 509-521.

Simons, H. (1980). *Towards a Science of the Singular*, Norwich, University of East Anglia: CARE Occasional Publications

Stake, R.E. (1967). *The countenance of educational evaluation* in Teachers College Record, 68:7 pp. 523-540.

Stake, R.E. (2004). *Standards-Based and Responsive Evaluation*, Thousand Oaks, Cal: Sage

Weiss, C & Rein, M (1969). `The evaluation of broad-aims programmes: a cautionary case and a moral' in *Annals of the American Academy of Political and Social Science*, 385, pp. 133-142

VALUES-ENGAGED EVALUATION[1]

Jennifer C. Greene
University of Illinois at Urbana-Champaign

Introduction

This chapter presents the character of, and rationale for, the explicit naming and claiming of values in evaluation. The specific values advanced are the democratic values of equity and inclusion. Readers may wish to apply the logic of this argument to the values advanced by other legitimate approaches to evaluation.

Greetings

Greetings to all from Urbana, Illinois, USA – a small urban community surrounded by a majestic prairie with deep, rich soil on which Illinois farmers grow vast quantities of corn and soybeans. These crops are made into products ranging from food for animals to fuel for automobiles. Illinois farmers also grow and produce the blueberries and peaches, the fresh greens and broccoli, the bell peppers and onions, and the delicious goat cheese that I purchase at our Saturday morning farmers market, all summer long.

> *…It is indeed a privilege to live here and to benefit so much from the generosity of the land.*

> *…Yet, I know that not all community residents share the realities of my experience of privilege.*

I also have the privilege of living near, and working at, the highly regarded University of Illinois at Urbana-Champaign. This flagship university of the state of Illinois is home to wonderful scholars and educators of all kinds. Some snapshots of current activities at my university can be seen at www.illinois.edu

> *…It is indeed a privilege to be a part of this magnificent university and to benefit from its multiple resources and accomplishments.*

…Yet, I know that not all community residents share the realities of my experience of privilege.

* * * * * * * * * * * *

…Discontinuities in experiences of privilege … privileges not shared by all

…inequities in life's challenges and affordances

…the promises of democracy to redress these inequities

…evaluation in service to these democratic promises

…serving democracy through evaluation.

* * * * * * * * * * * *

In this chapter, I offer my views on the intersections of evaluation and democracy, as represented by a values-engaged approach to evaluation that is positioned directly in service to democratic values and ideals.[2] I present the conceptualization and justification for these ideas, alongside snapshots from evaluation practice. The conceptualization and justification focus on the nature and role of values in evaluation, particularly democratic values. The snapshots from practice endeavor to ground these ideas in the nitty-gritty challenges of evaluation on the ground.

Values and evaluation

Evaluation is the systematic collection, analysis, and interpretation of information about human phenomena (commonly, social and educational programmes) in order to make judgments about their quality and effectiveness – judgments which are then used for decision-making, accountability, improvement, critique, and social betterment, among other uses (definition adapted from Weiss, 1998). Evaluation is thus intrinsically judgmental, involving some criteria of "goodness" upon which judgments of quality and effectiveness are made. Evaluation is thus also intrinsically infused with values, because the selection of these criteria – alongside other evaluative decisions – rest on the privileging of some set of values over others. Yet, values are rarely explicitly named or advanced in any of our current portfolio of evaluation approaches (Greene, 2009 and 2011).

---

2   This values-engaged approach also has a significant "educative" component, through which evaluation endeavors to enhance diverse stakeholders' critical understandings of the programme being evaluated, and its associated value claims and stances (Greene, Boyce, & Ahn, 2011; Greene, DeStefano, Burgon, & Hall, 2006).

The current portfolio of evaluation approaches

The evaluation community is at present a community of plenty. We are rich with diverse ideas about how to approach our work, with what questions and rationale, in service of whose interests and agendas (Alkin, 2004). In my view, this diversity of evaluation approaches is a sign of a healthy and vital community.

Among the evaluation approaches of currency and common use today are the following:

- Utilization-focused evaluation, championed originally by Michael Patton and now many others, aims primarily to be of instrumental use to specific identified evaluation users. Use in this approach is mostly commonly programme improvement of relevance to programme managers (see Patton, 2008).

- Accountability-oriented evaluation, advanced as part of the New Public Management movement popular in governments around the world, focuses on meeting established *indicators or benchmarks* of satisfactory performance, and is most useful to onsite programme managers and offsite administrators, such as civil servants and bureaucrats (see Newcomer, 1997).

- Impact evaluation, advanced by economists and others in the service of the interests of policymakers, endeavors to measure intended programme outcomes and to establish strong causal claims about them (see the Center for Global Development website; Joint Conference of AFREA and 3ie, 2009; Leeuw, & Vaessen, 2009).

- Educative evaluation, advanced mostly by evaluators themselves, uses the lenses of programme theory and logic modeling to map the logic of the programme as designed, implemented, and experienced, and also to assess both the quality of these various maps and the logical defensibility of programme outcomes, both intended and not (see Donaldson, 2007; Pawson & Tilly, 1997; Rogers et al., 2000; Weiss, 1998; W.K. Kellogg Foundation, 2004).

Evaluation approaches that are an important part of our landscape, but used less often include:

- Democratic and participatory evaluations, which seek democratizing social change in the contexts being evaluated and, ideally, beyond (Greene, 2006; Greene, Millett, & Hopson, 2004; House & Howe, 1999; Karlsson Vestman & Segerholm, 2009; Whitmore, 1998).

- Responsive approaches that focus on deep contextual, cultural, and pluralistic programme understandings (Hood, Hopson, & Frierson, 2005; Stake, 2004; Thomas & Stevens, 2004).

- Critical evaluation that engages a critique of social structures and institutions as part of assessing programme design and quality (Everitt, 1996).

Interestingly, these less commonly-used evaluation approaches are explicitly values-engaged, even values-committed, in contrast to the more common approaches, in which values are characteristically masked and muted, using an ostensibly values-neutral stance (Schwandt, 1997). Yet, values are present in all of our work.

Where do values show up in evaluation?

Values are present in virtually all aspects of evaluation, and they show up visibly in what might be called the common places of evaluation theory and practice, or the constitutive elements of our work.[3] With respect to this significant presence of values in our work, the most salient of these evaluation common places are the following:

- evaluation's *purpose and audience*, and related to purpose and audience, the intended evaluation *use(s)*;

- key evaluation *questions* to be addressed;

- *criteria* used to make judgments of quality;

- *social relational aspects* of the evaluator's *role* in the study.

That is, values enter our evaluation spaces primarily through decisions about whose interests, key questions, and agendas should be addressed – those of our policy and decision-makers, those of the programme developers and administrators, those of the programme staff and implementers, those of the intended programme beneficiaries and their families, those of the general public or advocacy groups or taxpayers? Within the extant portfolio of evaluation approaches presented just above, there are approaches oriented around the interests of each of these various evaluation stakeholder groups, all of whom can lay legitimate claim to evaluation resources

---

3   One framework for these evaluation common places is offered by Shadish, Cook, & Leviton (1992). Appendix 1 presents an adaptation of this framework. See also Greene, Boyce, and Ahn (2011).

and results.[4] Commonly used evaluation approaches characteristically address the interests of those with responsibility and power – policymakers, administrators, programme developers. Less commonly used evaluation approaches characteristically address the interests of those with less power and voice even in the decisions that directly affect their own lives – programme staff, and especially intended programme beneficiaries, their families, and their communities. Values further penetrate evaluation practice through the social relationships established in the contexts and spaces in which we work – relationships of power and privilege, of professional boundaries and discipline, of caring and reciprocity, of trust and acceptance, and more (Greene, 2011).

Yet, evaluators are usually not explicit or transparent about the values being promoted in any given evaluation study. Few evaluators explicitly name the values that are being promoted in their work; few even explicitly articulate the criteria being used for making judgments of programme quality. I believe that evaluators need to be far more attentive to the value dimensions of their own practice. We need to identify, name, and justify these values – both for ourselves and especially for those in the contexts in which we are working. I realize this is not standard practice, but I think it should be. In some ways, it is ironic that a practice called eVALUation is silent about the values being invoked in any particular study.

Democratic values and evaluation

Trained many years ago in traditional evaluation theories and objectivist methodologies, my current stance of embracing the inevitable presence of values in evaluation has come about through a journey. Major signposts along this journey for me included learning about evaluation from such grand masters as Lee Cronbach and Egon Guba; experiencing the heady excitement of the great qualitative-quantitative debate in all of its philosophical, methodological, and especially political glory; and continuing to anchor my own thinking about evaluation in my evaluation practice, thus privileging the realities of those our evaluation craft endeavors to serve.

---

4   Indeed, one of the greatest challenges of evaluation practice is making justifiable decisions about whose interests will be served and which values will be advanced in an evaluation study. Yes, some evaluation purposes and audiences are decided ahead of time, given to us as evaluators with little space for negotiation. BUT, in most evaluation contexts, there is also some *discretionary space* in which you as the evaluator can and should assert your own evaluation ideas, your own mini-theory of your craft – which is most importantly, your own ideas about whose interests should be served and thereby what values should be advanced by your work. These decisions are part of the evaluator's professional responsibilities.

The case for democratic values in evaluation

Once I fully accepted the presence of values in evaluation, I could identify no more defensible values than the ideals of democracy, equity, social justice, equality, freedom. In this, my thinking was influenced both by my own coming of age in the turbulent social-political movements of the 1960s in the US and by the inspiring work of my evaluation predecessors, especially fellow American Ernest House.

> *"Evaluation always exists within some authority structure; some particular social system. It does not stand alone as simply a logic or a methodology, free of time and space, and it is certainly not free of values or interests. Rather, evaluation practices are firmly embedded in and inextricably tied to particular social and institutional structures and practices."*
> House & Howe, 2000, p. 3

Evaluation, that is, is both constituted by the existing social and political order and, in turn, serves either to reinforce the existing order *or* to challenge and reconstitute it in particular ways. So, given that evaluation is embedded in and constitutive of the fabric of public decision-making, evaluation "should be explicitly democratic" (House & Howe, 2000, p. 4). More specifically, House and Howe's deliberative democratic evaluation approach focuses on insuring that the interests of *all* legitimate stakeholders are included, specifically those of the powerless and the poor, or those stakeholders whose interests are usually excluded from evaluative deliberations and decisions. In this way, deliberative democratic evaluation advances "an egalitarian … conception of justice that seeks to equalize power in arriving at evaluative conclusions" regarding effective social programmes and policies (House & Howe, 1999, p. 134).

> *"It would not be right for evaluators to provide evaluations only to the most powerful or to sell them to the highest bidders for their own uses … The interests of all stakeholder groups are central [to evaluation], and the interests of all relevant parties should be represented, as genuine democracy would require.... Evaluators must design evaluations so that relevant interests are represented and so that there is some balance of power among them, which often means representing the interests of those who might be excluded in the discussion, because their interests are likely to be overlooked in their absence."*
> House & Howe, 1999, p. 98

Democratic values-engagement in evaluation

With the significant influences of Ernest House and others, I have been developing and field testing with colleagues a *values-engaged, educative approach to evaluation* over the past several years. This section provides an overview of the values-engaged strands of our ideas.

The term "values-engagement" is intended to signal explicit attention to values as part of the evaluation process and to the central role that values play in our evaluation practice. From the framing of evaluation questions to the development of an evaluation design and methods, and from the interactions of stakeholders in the evaluation process to the especially important task of making judgments of programme quality, values are centrally featured in this approach. Engagement thereby suggests a kind of quiet insistence that questions of value be addressed throughout the evaluation, at every turn and every decision point – so values become interlaced with, knitted and knotted within evaluative thinking and judging. This is our aspiration.

The values-engaged approach both *describes* the plurality of values that exist in the evaluation programme and context and *prescribes* particular values – those of *democratic inclusion and equity*. Our emphasis on describing multiple and diverse values is well grounded in the pluralistic responsive evaluation work of Robert Stake (Stake 2004). And, as noted, our commitments to foregrounding democratic values are anchored in evaluation's proud democratic tradition (see Greene, 2006). Moreover, while descriptive approaches to values have the advantage of being "more politically and socially practical in a system of pluralistic interests," "prescriptive theories give evaluators a critical perspective and intellectual authority that descriptive theories cannot match" (Shadish et al., 1992, pp. 49-50).

Specifically, we first aim (ideally) to *inclusively describe and engage* the interests, perspectives, and values of *all* legitimate stakeholders in our evaluation, with particular attention to ensuring inclusion of the interests, perspectives, and values of those least empowered and traditionally not heard in that context. The interests of the majority are not excluded in this approach; rather the interests of the minorities are specifically and intentionally included (House & Howe, 1999). To illustrate our attention to values, in the context of an evaluation of a science education programme, some stakeholders may prefer "teacher-directed instruction" because they believe

it can most effectively enhance student mastery of science content knowledge, which is the outcome they value most highly. Other stakeholders may support "problem-based learning" because they believe it advances students' scientific reasoning skills and motivation to learn more science, which are the outcomes that they value most highly. These stakeholders' instructional viewpoints and their accompanying value stances would be surfaced in our values-engaged approach – all with respect and acceptance – for engagement, dialogue, and discussion.

Stakeholder inclusion has a long history in evaluation, from both utilization and democratic traditions. Our emphasis is on the inclusion of all voices and values, precisely because such inclusion is more pluralistic, more equitable, and more just. We believe that by actively seeking to include, respect, and represent the legitimate plurality of stakeholder interests and values, evaluation itself can increase awareness of the importance and acceptance of the intrinsic diversity of experience and perspective in the programme being evaluated, and thereby the diversity of values and beliefs that accompany programme experiences and their meaningfulness (Greene et al., 2006).

Beyond descriptive inclusion, our evaluation approach also seeks to prescribe evaluative *engagement with values of equity*. We define equity as being concerned with the *treatment* of all programme stakeholders. Treatment refers to access and opportunity to participate in and to benefit from a programme. An equitable programme offers all participants equal *access* to the programme, and equity in the character of both programme *experiences* and *accomplishments*. For example, do all potential learners in an educational programme have an equal opportunity to be recruited or selected for participation? Is the pedagogy used and content offered in the programme appropriate and meaningful for all participating learners? Do all learners have equitable opportunities to achieve and excel in the programme? Our commitment to equity is enacted through generating evaluation questions, data, and discussions related to the ways in which a programme is attending appropriately and with meaningful consequence to *all* individuals and groups that are present in the context, particularly those that have been identified as being traditionally underserved.

To illustrate, consider an evaluation of a technology education programme designed to enhance the individual and collective "hightech" capacity of rural residents and their villages in selected Cen-

tral American countries. The programme is funded by the Gates
Foundation in the US. Here (box 1) is one possible democratic val-
ues-engaged statement of evaluative intent in this context.

Box 1: Example of democratic values-engaged statement

This evaluation is designed to concentrate on the ways in which, and the extent to
which, this programme provides equitable access and opportunity-to-learn, for the most
disadvantaged and isolated of rural residents and their villages in the countries being
served. The evaluation will assess implementation quality and outcome attainment
– both intended and unintended – for all participants, but again concentrate on
those most under-served. The evaluation will endeavor to inclusively represent the
perspectives and experiences of the programme's funder, the Gates Foundation, and
programme administrators and staff, all in respectful dialogue with one another and
with the diverse stances and experiences of programme beneficiaries, their families, and
their communities. The evaluation aspires to enhance deep programme understanding,
specifically the diversity of experiences and meaningfulness of programme
participation, on the part of all stakeholders, and further to catalyze programme
changes, as needed, that serve the under-served more equitably and successfully.

Democratic Values-engaged evaluation in practice

This section presents illustrative guidelines for the Values-engaged
evaluation practitioner, with examples offered (box 2) for a hypo-
thetical programme evaluation context (see also Greene, Boyce, &
Ahn, 2011, for additional practical guidance).

Box 2: Hypothetical programme evaluation context

Access to the extraordinary resources of the *World Wide Web* (WWW) is not yet world-
wide. A significant fraction of the populations of countries in the "developing" world
lack this access. In a ground-breaking public-private partnership, the European Union
and Google have teamed up to develop both the infrastructure and human resources
needed to bring the internet to people and places who do not yet have access to the
information of the contemporary global knowledge economy.

One set of programme initiatives in this grand WWW partnership involves the training
of people in places lacking internet access, in the technical skills needed to develop,
maintain, and purposefully use web access for local constructive development agendas.
The first cluster of people and places to be trained in this programme is located in sub-
Saharan Africa.

Illustrations of practical guidelines for democratic Values-engaged evaluation are presented using some of the evaluation common places offered by Shadish et al. (1992) (see also Appendix 1).

1. *What are relevant characteristics of the programme to be evaluated and its (policy and practice) context?*

In describing the programme-in-context, the evaluator attends to:

- The contextual meanings of diversity, or what dimensions of diversity and difference matter most in the context at hand and how those dimensions are defined or understood in context.[5] For the WWW programme, education and income are likely candidates for macro diversity differentiations. But, what else matters to individuals' sense of place and hope in these sub-Saharan African communities? Strategic interviews, informal observations, reading newspapers and other documents would be among the strategies useful to develop these contextual understandings.

- Understandings of the programme's theory or logic, along with accompanying value commitments, from the perspectives of multiple stakeholders. A programme theory for this context could address the following kinds of questions. In the WWW training project, what is the depth and currency of the technology content being provided? Is the pedagogy being used appropriate for the learners involved and does it incorporate relevant research (e.g., on culturally appropriate pedagogy) in its design and implementation? Who is recruited for this project and why? Who might be left out? Is the logic connecting activities and outcomes strong, defensible, and supported by existing research? Programme theories could be generated via concept mapping or interviews with multiple groups of key stakeholders. They then serve as a vehicle for critical stakeholder engagement with their own and others' underlying programme assumptions, values, and aspirations.

2. *What are the primary purposes of and audiences for the evalu*ation?

The primary purposes of democratic Values-engaged evaluation are (a) to generate a deep contextualized understanding of the WWW training programme, as designed and as experienced, on the part

---

5   Diversity is understood to include historical socially-constructed characteristics of difference and disadvantage, as well as the myriad other ways in which humans are different one from another (Greene, Boyce, & Ahn, 2011).

of diverse stakeholders, and (b) to advance the interests and well-being of those least well served in the communities involved in the WWW programme.

Evaluation in this approach is oriented around the interests of four key stakeholder audiences, in order:

• programme leaders and staff, as well as community leaders and staff in the sites where the programme is implemented;

• programme participants, their families, and other community members;

• programme decision and policymakers (local and distant), including the EU and Google; and

• the broader communities of technology education researchers and educators.

3. *What are the key evaluation questions?*

Key evaluation questions in this approach could include the following:

• What is the quality and match of the WWW technical training programme content and pedagogy, as designed and as experienced, to the characteristics of the learners being served?

• What is the contextual power, relevance, and meaningfulness of the programme as designed and as experienced for the learners involved, or how well and with what consequence does the programme 'show up' in their lives (Kushner, 2000)?

• What is the quality, magnitude, and contextual importance of programme outcomes, both intended and unintended?

• In what ways and to what extent does the programme advance the interests and well-being of those least well served in the WWW technical training programme contexts?

Specific evaluation questions are developed collaboratively with key stakeholders.

4. *By what criteria will programme quality be judged?*

A good WWW technical training programme:

• Has current and relevant technology content, and contextually and culturally appropriate pedagogy – according to relevant educational standards, the views of key stakeholders, and relevant research;

- Is one that is meaningful and relevant to the lives of participants in some possibly sustaining way;

- Is one in which the most underserved in the communities involved have meaningful access to the programme, high quality programme experiences, and accomplishments on a par with their peers.

5. *What methodological framework and accompanying assumptions are to be used?*

Most consonant with the values and stances of this approach is a mixed methods framework, in which the mixing happens at *all* levels of philosophy/paradigm, methodology, and method. Especially important is the mixing at the level of philosophy/paradigm because such a mix requires a legitimization of and respect for different ways of seeing and knowing. Such a mix also invokes a dialogic engagement with 'difference', toward 'listening well' and understanding of the 'other.' A methodological framework that itself engages with 'difference' can well support and even reinforce the values of inclusion and equity.

6. *What are appropriate evaluation design features and methods?*

Values-engaged evaluation is not distinctively about method. Within a dialogic mixed methods framework, many different methods and methodologies will accomplish the agenda of this approach. One could further envision conducting a "fleet of smaller evaluation studies" (Cronbach & Associates, 1980), rather than one grand study, as this would be highly consonant with the stances of democratic pluralism and inclusion.

7. *What is the appropriate evaluator role?*

Broadly, the evaluator role is critical in this approach:

- The evaluator must have contextual authority and credibility – either pre-established or earned at the outset of the evaluation;

- The evaluator must focus as much on the 'social relations' of the evaluation study as on its technical and data-oriented tasks and content. These social relations importantly help to constitute the information/knowledge generated;

- The evaluator is at root an educator;

- The evaluator's work is engaged primarily via dialogue.

*8. How will communications with key stakeholders be conducted?
How will evaluation findings be reported and disseminated?*

Ongoing, frequent communications and reports will be offered to
multiple WWW programme stakeholders via a variety of formats
(written; oral; performative; web-based; narrative), along with
efforts to engage various stakeholders in ongoing dialogues about
programme quality and equity. The varied formats are intended to
evoke deeper stakeholder engagement with key issues at hand,
especially with challenging and sensitive issues.

And so …

May we join together in encouraging evaluators around the world
to be explicit about the values they are advancing in their evaluation
practice, and to seriously consider positioning their evaluation work
in service to democracy.

References

Alkin, M.C. (2004). *Evaluation roots: Tracing theorists' views and influences.* Thousand
Oaks CA: Sage.

Center for Global Development, references to 'impact evaluation' for international
development initiatives at: http://www.cgdev.org/content/search/?search_
query=impact+evaluation&x=2&y=4

Cronbach, L.J., & Associates (1980). *Toward reform of programme evaluation.* San
Francisco: Jossey-Bass.

Donaldson, S.I. (2007). *Programme-theory driven evaluation science: Strategies and
applications.* New York: Routledge.

Everitt, A. (1996). 'Developing critical evaluation'. *Evaluation*, 2(2), 173-188.

Greene, J.C. (2006). 'Evaluation, democracy, and social change'. In I.F. Shaw, J.C.
Greene, and M.M. Mark (eds.), *The SAGE handbook of evaluation* (pp. 118-140). London:
Sage.

Greene, J.C. (2009). 'Values in evaluation: Revelation and responsibility.' Invited keynote
address presented at the Second South African Monitoring and Evaluation Association
Conference, August, 2009, Johannesburg South Africa.

Greene, J.C. (2011). *Valuing evaluation's values.* Plenary address presented at the annual
conference of the American Evaluation Association, November 2011, Anaheim CA.

Greene, J.C., Boyce, A., & Ahn, J. (2011). *A values-engaged, educative approach for
evaluating educational programmes: A guidebook for practice.* University of Illinois.
(available in the American Evaluation Association e-library, http://www.eval.org)

Greene, J.C., DeStefano, L., Burgon, H., & Hall, J. (2006). 'An educative, values-engaged approach to evaluating STEM educational programmes'. In D. Huffman and F. Lawrenz (eds.), *Critical issues in STEM evaluation. New Directions for Evaluation 109* (pp.53-71). San Francisco: Jossey-Bass.

Greene, J.C., Millett, R., & Hopson, R.K. (2004). 'Evaluation as a democratizing practice. In M. Braverman, N.K. Constantine, &J.K. Slater (eds.), *Foundations and evaluations: Contexts and practices for effective philanthropy.* San Francisco: Jossey-Bass.

Hood, S., Hopson, R., & Frierson, H. (2005). *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice.* Greenwich CT: Information Age.

House, E.R., & Howe, K.R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

House, E.R., & Howe, K.R. (2000). 'Deliberative democratic evaluation'. In K.E. Ryan & L. DeStefano (eds.), *Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation. New Directions for Evaluation 85* (pp. 3-12). San Francisco: Jossey-Bass.

Joint Conference of AFREA and 3ie (2009) on impact evaluation.http://www.3ieimpact. org/admin/pdfs_papers/50.pdf

Karlsson Vestman, O., & Segerholm, C. (2009). 'Dialogue, deliberation, and democracy in educational evaluation'. In K.E. Ryan & J.B Cousins (eds.), *The SAGE international handbook of educational evaluation* (pp.465-482). Thousand Oaks CA: Sage.

Kushner, S. (2000). *Personalizing evaluation.* London: Sage.

Leeuw, F., & Vaessen, J. (April 2009). *Impact evaluation and development. NONIE guidance on impact evaluation.* (April 2009). Developed for NONIE, the Network of Networks on Impact Evaluation. http://www.worldbank.org/ieg/nonie/docs/Guidance_ IE.pdf

Newcomer, K.E. (ed.) (1997). 'Using performance measurement to improve public and nonprofit programmes'. *New Directions for Evaluation 75.* San Francisco: Jossey-Bass.

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

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London: Sage.

Rogers, P.J., Hacsi, T.A., Petrosino, A., & Huebner, T.A. (eds.) (2000). 'Programme theory in evaluation: Challenges and opportunities'. *New Directions for Evaluation vol 87.* San Francisco: Jossey-Bass.

Schwandt, T.A. (1997). 'Reading the 'problem of evaluation' in social inquiry'. *Qualitative Inquiry,* 3(1), 4-25.

Shadish, W.R., Jr., Cook, T.D., & Leviton, L.C. (1992). *Foundations of programme evaluation: Theories of Practice.* Thousand Oaks CA: Sage.

Stake, R.E. (2004). *Standards-based and responsive evaluation.* Thousand Oaks CA: Sage.

Thomas, V.G., & Stevens, F.I. (eds.) (2004). 'Co-constructing a contextually responsive evaluation framework'. *New Directions for Evaluation 101.* San Francisco: Jossey-Bass.

Weiss, C.H. (1998). *Evaluation, second edition.* Upper Saddle River, NJ: Prentice Hall.

Whitmore, E. (ed.) (1998). 'Understanding and practicing participatory evaluation'. *New Directions for Evaluation no 90*. San Francisco: Jossey-Bass.

W.K. Kellogg Foundation, *Logic Model Development Guide*. http://www.wkkf.org/knowledge-center/resources/2010/Logic-Model-Development-Guide.aspx

Appendix 1

Evaluation Common places (adapted from Shadish, Cook, & Leviton, 1992)

What is the nature of meaningful and societally valued educational change?

1. What are relevant characteristics of the programme to be evaluated and its (policy and practice) context?

2. What are important features of the evaluation context?

What is the role of values in evaluation?

3. What are the primary purposes of and audiences for the evaluation?

4. What are the key evaluation questions?

5. By what criteria will programme quality be judged?

What is the nature of warranted evaluative knowledge?

6. What methodological framework and accompanying assumptions are to be used?

7. What will constitute evaluation quality or success?

What is defensible evaluation methodology?

8. What are appropriate evaluation design features and methods?

9. What is the appropriate evaluator role?

What is meaningful and consequential evaluation use?

10. How will communications with key stakeholders be conducted? What are important forms of process use?

11. How will evaluation findings be reported and disseminated?

Part 3
Examples
of Equity-focused
evaluations

EVALUATING THE CONTRIBUTION OF UNDP TO EQUITY-FOCUSED PUBLIC POLICIES IN BRAZIL AND CHINA

Oscar A. Garcia, Evaluation Adviser, UNDP Evaluation Office
Juha I. Uitto, Deputy Director, UNDP Evaluation Office

Introduction

Evaluating the performance of public policy is a fundamental ingredient to foster accountability, good governance, and to improve development effectiveness. Distinct from the commonly used term 'aid effectiveness' which focuses on the external assistance given to a particular country from a donor's point of view, development effectiveness takes the national needs and aspirations as its starting point. In China and Brazil, two emerging powerhouses, it has a very specific meaning, with their many development challenges remaining despite remarkable economic growth. In China, public policy performance evaluation took the form of the Xioakang society, a society characterized by a focus on 'five balances' – between urban and rural; between different geographical regions; between economic and social; between people and nature; and between domestic development and opening-up beyond China's borders. In Brazil it took the form of 'paying the social debt' which meant going beyond economic growth and focusing on the quantity and quality of social expenditure and more importantly reducing the gap between the immense majority of Brazilians and the few who are well off.

The United Nations Development Programme (UNDP) Evaluation Office undertakes country- level evaluations, called Assessment of Development Results (ADR), to assess the contribution made by the organization to the achievement of national development results, as defined by governmental plans and priorities. The ADR is a distinctive type of country-level evaluation. It focuses on assessing the contribution made by UNDP, to the achievement of development results in a certain country, usually over a period of eight years, or two programming cycles. Additionally the ADR tries to capture the strategic positioning of UNDP as a development partner, with the understanding that UNDP is but one among several partners that contribute to development. Assessing the way UNDP responds to

changes in the development context of each country and how it establishes partnerships to foster sustainable human development, is of crucial importance to evaluate its relevance and effectiveness. The ADR is also useful for strengthening the accountability of UNDP towards member states and towards its governing, body and it will also contribute to organizational learning.

This paper is an attempt to extract lessons from evaluative evidence gathered from the ADRs conducted by the UNDP Evaluation Office in China and Brazil (EO 2010, 2011). The paper focuses on the role played by international cooperation, particularly by UNDP and other international partners, in support of equity-focused public policies. The first part provides a brief overview of historical trends in inequality in Brazil and China. The second part then outlines the main findings of the evaluations regarding UNDP contributions towards policies that address inequalities in the two countries. The paper ends with a brief section on lessons learned and conclusions.

Public policies focused on equity

Brazil and China have both gone through a period of impressive economic growth, which has not benefited equally all segments of the societies. Both have identified public policies to reduce poverty and inequalities. The two countries have some similarities. Each has the largest population in their region, China with 1.3 billion people and Brazil with 194 million people. Both are the largest economies within their respective regions: China in Asia with a GDP of US$4,985 billion and Brazil in Latin America and the Caribbean with a GDP of US$1,594 billion as of 2009[1]. They also host a large number of poor people. In Brazil, 21.4 per cent of people in 2009 were below the national poverty line representing nearly 40 million people (down from 30.8 per cent in 2005). In China, just 2.8 per cent of people were counted to be below the national poverty line in 2004. However, given the huge population size, this figure translates into some 37 million people.
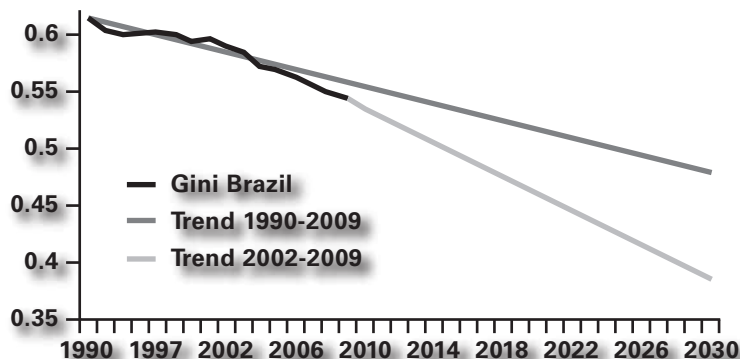
Brazil

Inequalities in Brazil have been reduced in the past twenty years. Inequality, as measured by the Gini coefficient, has decreased from 0.614 in 1990 to 0.543 in 2009.

---

1   World Bank data (http://data.worldbank.org/).

In 1988 a new constitution was drafted in Brazil, which transformed many public policies, particularly the social policies. These are rights-based policies and include the participation of citizens in their design, implementation and oversight.

Figure 1: Gini coefficient of Brazil, from 1990 projected to 2030

Source: IPEADATA and World Bank Database

High economic growth came with a transformation of the living conditions of the Brazilian population when it led to income redistribution. The Gini coefficient, that measures the evolution of income distribution, showed a systematic decline in the last twenty years. However it still shows high levels of inequality that hamper the capabilities of the majority of the Brazilian population. While in 1990 the proportion of families below the poverty line was 42 per cent, in 2009 that proportion was close to 21 per cent. Similarly the proportion of families in extreme poverty fell from 20 per cent to 7 per cent of the population[2].

The reduction of poverty and inequalities cannot be interpreted merely as the result of impressive economic growth; it is also the result of expanding the scope of the social policies and increased public investment in the social sector. In 2010 the public expenditure in social sector reached 9.8 per cent of GDP, which is much higher than in many countries with similar development (IPEA 2010)[3].

---

2   Based on the national household surveys (Pnad/IBGE). See www.ipeadata.gov.br

3   Gastos com a Política Social: alavanca para o crescimento com distribuição de renda. Brasília: Comunicados IPEA, N° 75.
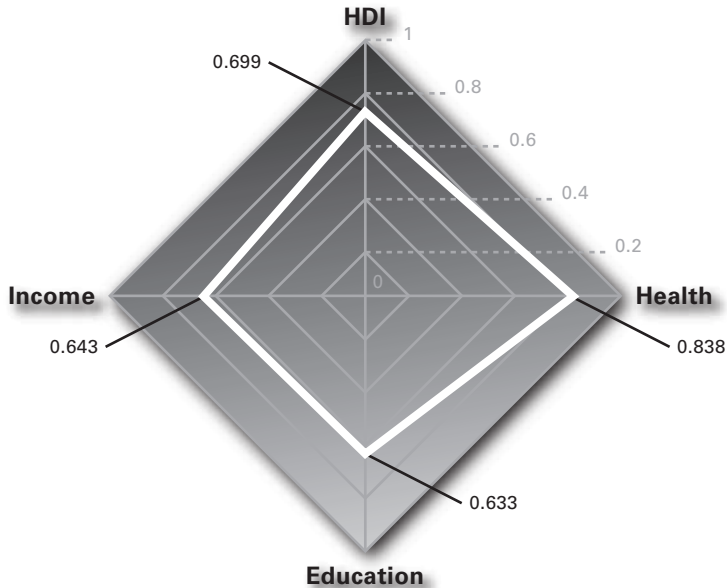
In Brazil these changes were the result of a systematic and con-certed effort, made by at least three Brazilian administrations, to reduce the gap that divides the haves from the have-nots. Brazil launched a well-known conditional cash transfer programme, Bolsa Familia, which covers 12.4 million families below the poverty line. Again, however, the results cannot be explained by the success of such conditional cash transfer programmes alone. It is also the result of universal access to health or education, and the role played by the 'Unified National Health System', which offers free cover-age of health services to the poor and improved the life expectancy at birth from 63.2 years in 1997 to 69.3 years in 2008[4]. The 1988 constitution declared health care to be the right of the citizen and its provision the duty of the state. The unified system was created in 1989 from the merger of two state systems, one for those in formal work and the other for everyone else. The percentage of the population with access to sanitation has also increased, from 45 per cent in 2001, to 53 per cent in 2009. It is not only the deliberate increase of the national minimum wage (which between 1995 and 2010 increased in real terms by more than 100 per cent), but a com-bination of all of the above plus programmes targeted at vulnerable populations. These include the rural old and people with disabilities, who now receive retirement benefits even if they have not contrib-uted previously to retirement plans. Employment programmes have also been established for young black males who are victims of sys-tematic violence and insecurity.

The above facts indicate that reduction of inequalities in Brazil is the result of a combination of public policies addressing social protec-tion and income distribution, as it is the multidimensional approach taken by successive Brazilian governments that has reduced pov-erty and inequalities. This approach resonates with the human development paradigm. The paper does not pretend to show any causality, and most probably the contribution made by UNDP was minimal, but the concept resonates with what has been advocated by the Human Development Reports since 1990.

Figure 2, below, illustrates the progress made in the reduction of inequalities, as measured by the Human Development Report according to the new methodology adopted in 2010 (HDR 2010).

---

4   See (datasus.gov.br)

Figure 2: Distribution of HDI components in Brazil, 2010

According to these new criteria the component in which Brazil is best positioned is health, whereas education lags behind since the new measurement includes not only the enrolment rates and literacy levels, but also the gap between the average years of schooling and the expected years of schooling.

Regional disparities also remain high. Inequalities in welfare and poverty between and within regions are stark, particularly between the lagging North-east region and the South-east region and also between urban and rural areas. In a country with a vast territory, the response was the decentralization of social policies to sub national levels. Since 1988 when a new constitution was promulgated, different approaches to effectively decentralize the implementation of social policies have been tested. Special attention was given to the North-east region. However, the results are mixed as the institutional capacity to implement the social policies, particularly at municipal level, varied greatly.

Inequality in Brazil still remains inexcusably high. Despite the success of recent policies a sustained effort needs to be made to obtain better results. The debate is shifting from the amount of resources deployed to social policies to the quality of these policies and the improved provision of social services. For that purpose the evaluation of public policies is considered as a necessary instrument for the achievement of development effectiveness.

China

China has experienced the most spectacular growth of any country in the world over the past twenty years. However, the accelerated growth has also increased the income inequality and led to widening disparities between people and regions. In China, inequalities have a distinctly regional dimension. The coastal areas in the east and south of the country and their major cities – Beijing, Shanghai, Hong Kong and others – are thriving and have per capita income levels comparable and higher than many developed countries. At the same time, large parts of China, especially in the west, lag severely behind and retain the characteristics of poor developing countries as measured by social and economic development indicators. Despite huge strides in economic development, some 14.8 million people, or 1.6 per cent of the total rural population in 2007 lived in absolute rural poverty, according to the government's poverty line. In addition, many migrant workers who have left the countryside for jobs in the booming cities are highly vulnerable and lack social protection.

Income gaps between urban and rural areas and between regions have widened and continue to do so. The Gini coefficient has deteriorated from 0.20 to 0.40. Table 1 below illustrates how the inequalities, although still low in comparison to Brazil, have increased (HDR 2010). There is also a strong social component, which manifests itself clearly in health statistics. Maternal mortality is as much as seven times as high in some remote areas as in the eastern coastal region, which reflects disparities in factors such as safe water and hygiene.

Table 1. Human development indices of China, 1975-2010

Country	Year	Income	Education	Health	Overall
China	1975	0.16270	0.34540	0.30202	0.27406
China	1980	0.15584	0.30620	0.26825	0.24605
China	1985	0.11807	0.28950	0.23984	0.21903
China	1990	0.17682	0.27540	0.21468	0.22337
China	1995	0.22562	0.21330	0.19355	0.21094
China	2000	0.22487	0.15390	0.17483	0.18508
China	2005	0.26599	0.12790	0.15894	0.18649
China	2010	0.27010	0.10370	0.14620	0.17645
Within Country Inequality	1975-2010	66.00%	-69.98%	-51.59%	-35.62%

*Note: Within country inequality based on the Atkinson inequality index with inequality risk aversion coefficient
equal to 1*

*Source: HDRO calculations based on Sala-i-Martin and Pinkovskiy data (for income), Barro-Lee data
(for education) and UN Population Division (for health/Life Tables)*

Added to this is the severe environmental pollution and degradation
of natural resources that has accompanied the unchecked growth,
industrialization and urbanization. Some analysts argue that environ-
ment will, in the future, be the critical factor setting limits to eco-
nomic development in China (Economy 2004). On the social front,
pollution is taking a heavy toll on the health of the population (Lu
and Gill 2007). Many of these problems, whether severe degrada-
tion of water and land resources in the rural areas or pollution in
industrial regions, affect the poor people most.

Recognizing these challenges, China has launched its vision of
Xiaokang society. 'Xiaokang' is defined as "all round, balanced and
harmonious society". In 2003, the *Xiaokang* vision was redefined to
emphasize a "Scientific Concept of Development' that focuses on
'five balances' – between urban and rural, between different geo-
graphical regions, between economic and social, between people
and nature, and between domestic development and opening-up
beyond China's borders." The Chinese government, in response,
today pays great attention to improving governance in a broad range
of areas. China has initiated reforms of the public sector to improve
the performance of the civil service, institutionalize management

and accountability structures, and to improve the provision of services to the public.

What has been the contribution made by UNDP?

What has the UNDP contributed to the reduction of inequalities? Perhaps it is an unfair question but, nevertheless, international cooperation has played a role in supporting the development aspirations of countries. The ADRs have tried to answer the question as it pertains to UNDP and its role. The evaluations further attempt to identify what has and what has not worked, how can UNDP's support be improved, and can we learn anything from the experience.

The ADRs use mixed methods to collect and analyze evidence. By virtue of the nature of UNDP's work, there is an emphasis on qualitative methods. Triangulation is used to simultaneously validate data and information collected from different sources, including surveys, interviews, document analysis and national statistics.

Brazil

The evaluation found that UNDP was relevant to Brazil's national aspirations either through advocacy initiatives or through the implementation of projects contributing to the achievement of development results. It made contributions through advancing the human development concept in Brazil and also, more specifically, UNDP was instrumental in the implementation of the conditional cash transfer programme, Bolsa Familia, through the establishment of a national security policy with an integral approach that put emphasis on prevention. It was instrumental in the fiscal adjustments of many states; it supported the institutionalization of environmental policies and it was instrumental in the implementation of South-South cooperation as part of the Brazilian foreign policy. UNDP has also supported national efforts to strengthen the institutional capacity for implementation of social policies in health, education and housing sectors.

The impact of the national human development reports and the introduction of human development indexes at municipal level, have helped measure the regional and local disparities, and contributed to focusing attention on the important income and regional disparities that existed. It has also advocated addressing sensitive issues, such as gender equality and racism as part of public policies.

The evaluation found that UNDP made strategic contributions to the formulation of social policies although its strategic relevance has diminished in recent years.

The evaluation put particular emphasis on the sustainability of UNDP's results. It found that, while UNDP's contribution has been significant for the conceptualization of several policies, its role is no longer predominant, because positive changes in the institutional framework of the Brazilian state took place during the years covered by the evaluation. The impressive results, in terms of the public policies implemented to promote sustainable growth while reducing poverty and inequalities, show that the federal state gained more capacity. The support provided by UNDP in strengthening the capacity of national institutions in the environmental, health or education sectors was instrumental to that end. UNDP played a catalytic role in the conceptualization and implementation of some of these policies and the national institutions made further conceptual and operational improvements.

The benefits of UNDP interventions and the likelihood of their continuation over time is guaranteed, as these are national policies embedded into the institutional framework of the Brazilian state. It is worth noting that the role of the state has changed and was strengthen during the last two administrations of President Luis Inácio Lula da Silva. UNDP played a technical assistance role and in some cases, such as the elaboration of national human development reports, it also played an advocacy role. It should not extend its contribution beyond that. As identified before, there are institutional challenges for the implementation of public policies at the sub-national level. UNDP can provide technical assistance at the decentralized level only if it is so requested by the federal government and has the resources to do that properly.

China

UNDP was the first international organization to enter China in 1979 when the country was opening up. Since then, there has been a close partnership between the Chinese government and UNDP.

When the ADR was designed by the Evaluation Office, Chinese partners requested the addition of 'social equality' as a specific evaluation criterion to supplement the standard criteria of relevance, effectiveness, efficiency, responsiveness and sustainability. The evaluation consequently focused explicitly on answering whether:

- UNDP interventions and programmes contributed to reduced vulnerabilities in the country (regarding vulnerable groups, gender equality and regional disparities);

- they in any way influenced existing inequities in the society; and,

- the selection of geographical areas for UNDP's interventions was guided by need.

Overall, the stakeholders interviewed for the evaluation regarded UNDP's programme to be highly relevant to the country's development needs. This is largely due to the significant overlap between the Chinese development priorities and UNDP's planned outcomes, which have been well aligned with the 5-year plans. In particular in three strategic areas UNDP's contributions are acknowledged. First is the alignment of the Xiaokang vision with the Millennium Development Goals. Secondly, UNDP emphasis on gender equality is widely recognized as having emphasized mainstreaming gender concerns in various programmes. This has been the result of continuous advocacy on the part of UNDP. The third contribution derives from the special attention given by UNDP to vulnerable and disadvantaged groups in the areas of poverty reduction, democratic governance, and HIV/AIDS. The evaluation also found that UNDP has been adequately responsive to changing conditions and priorities in China.

As for actual development results and UNDP's contributions, the country programme was explicit about issues pertaining to inequality on various levels. The evaluation found that UNDP has effectively supported the strengthening of the rule of law to protect the human rights of all, especially poor and disadvantaged groups. Several key projects directly targeted vulnerable and disadvantaged groups such as migrant workers, rural farmers and HIV-affected people. Other projects included indirect benefits to vulnerable groups. UNDP typically identifies needs and makes a conscious effort to address the needs of vulnerable groups through its projects. Such efforts extended to post-disaster relief interventions, as in the case of the Sichuan earthquake that struck during the period under evaluation, in May 2008.

The sustainability of UNDP initiated efforts in China looks very promising because of the high degree of national ownership of the programme and its alignment with national policies and structures. The possible challenge may be in the environmental sector, especially in biodiversity conservation, where it is important to ensure continued funding from national sources. Environmental sustainability depends to a great extent on China's ability to harness a critical mass of environmental awareness among decision-makers and the public at all levels, and to translate this awareness into action.

As for the regional dimension, the ADR found that, while UNDP has established strong partnerships with the central government, the partnerships with provincial governments have been pursued less systematically. Consequently, engagement at the provincial level has remained relatively weak and should be strengthened. Similarly, partnerships with the private sector and civil society are still emerging, although rapid progress has been made in the recent years.

Lessons and conclusions

Evaluating policy impact of external assistance is challenging. After all, policies are the purview of the governments of the countries, not of international organizations or donor agencies. In large advanced developing countries such as Brazil and China the institutional structures and policy-making processes are complex and anchored in national political choices. Organizations such as UNDP play a role in supporting the policy processes through capacity development, dialogue and advice, as well as by highlighting successful examples from international experience.

The ADRs, as country level evaluations, are not expected to identify or establish the attribution of any of the national policy results to UNDP. They merely try to identify, in a plausible way, the contribution made by the organization to a policy formulation process that involves many actors at different stages. The contribution is easier to identify in the case of projects that lead directly to the formulation of public policies. It is more difficult to establish the link when the policies are influenced by a long lasting advocacy campaign, such as undertaken through the Human Development Reports. However, in the cases highlighted in this paper the external evaluators were confident enough to identify the contributions made in both cases.

In Brazil, the UNDP contribution to significant changes was found to be constant but mostly silent. At times, it was based on the advocacy capacity of the UN. In other instances, it was the result of a long lasting relationship between the government and UNDP in which there is room for experimentation. Catalytic change does not need amplifiers to make it happen. If certain experiments produce results that can be seen as successful, replication will often follow.

In China, regions identified for detailed case studies were selected reflecting their different levels of development. Attention was paid to issues beyond the level of economic development, to factors such as environmental vulnerability and energy problems. Analysis

of regional inequality and UNDP's response to it was thus built into the evaluation design. Furthermore, the evaluation team sought to identify UNDP contributions by carrying out content analyses of government policies in relation to what UNDP had advocated. Such techniques could establish a plausible contribution to policy formulation.

In conclusion it can be said that, in both cases, it was possible to establish with a certain degree of confidence, that UNDP had made important contributions towards policies that would reduce inequalities in Brazil and in China. Such contributions could not – and should not – be quantified, as the results themselves are national. Nevertheless, the support from UNDP has been valuable.

References

Economy, Elizabeth C., (2007). *The River Runs Black: The Environmental Challenge to China's Future.* Cornell University Press, Ithaca.

Institute of Applied Economic Research, (2010). Gastos com a Política Social: alavanca para o crescimento com distribuição de renda. *Comunicados IPEA*, N° 75. Brasília.

Lu, Xiaoqing and Bates, Gill, (2007). Assessing China's Response to the Challenge of Environmental Health. *China Environment Series* Issue 9, pp. 3-15. Woodrow Wilson International Centre for Scholars, Washington, DC.

UNDP, (2010). Assessment of Development Results: China – Evaluation of UNDP Contribution. UNDP Evaluation Office, New York.

UNDP, (2011). Assessment of Development Results: Brazil – Evaluation of UNDP Contribution. UNDP Evaluation Office, New York.

UNDP, (2010). The Real Wealth of Nations: Pathways to Human Development. *Human Development Report 2010.* UNDP and Palgrave Macmillan, New York.

USING A HUMAN RIGHTS APPROACH TO EVALUATE ILO'S DISCRIMINATION STRATEGY

Francisco L. Guzman,
Senior evaluation officer, ILO Evaluation Office

Introduction

In this article, I will summarize the challenges and lessons learned through the inclusion of the transformative paradigm in the methodology used for high-level evaluation of the International Labour Organization's (ILO) discrimination strategy. This implies an emphasis on accountability, strategic alignment, ownership, and organizational effectiveness in the implementation of the strategy, by focusing the assessment on:

(i)   the coherence[1] of the activities under the ILO discrimination strategic outcome with other country programme outcomes, that can contribute to the application of the ILO Declaration on Fundamental Principles and Rights at Work, and the International Human Rights (HR) and Gender Equality (GE) principles of non-discrimination, which are mutually reinforcing;

(ii)   the *relevance* of the ILO discrimination strategy to: the ILO's strategic and Programme and Budget (P&B) objectives and indicators, as well as to HR & GE, as defined by international and regional conventions; national policies and strategies; and, the needs of rights-holders and duty-bearers both women and men, who are targeted by an intervention;

(iii)   the *effectiveness* of the discrimination strategy, which involves assessing the way in which results were defined, monitored and achieved (or not) and how the processes that led to these results were aligned with HR & GE principles (e.g. inclusion, non-discrimination and accountability);

---

1   Including assessment of complementarities, coordination and consonance

(iv)   the *efficiency* in the implementation of the activities to ensure that implementation does not overtax available resources nor create insolvable bottlenecks;

(v)   the *impact* of ILO activities on constituents' efforts to strengthen national legal frameworks; and

(vi)   *sustainability* of enforcement systems that guard against discrimination, and the sustainability of both the final and the immediate impact of the results of the activities.

Evaluating the ILO strategy on discrimination through a human rights lens

A team consisting of a female external evaluator, a male senior evaluation officer from the ILOs evaluation section (EVAL), and three female research assistants conducted the evaluation of ILO's strategy on discrimination.

Evaluation approach

Mixing qualitative and quantitative approaches, while ensuring the inclusion of different stakeholders (such as groups most vulnerable to discrimination), offered a wide variety of perspectives and a more reliable picture of reality. This, however, does not mean that large quantitative studies cannot benefit from HR & GE analysis. On the contrary, every evaluation has the potential to assess these areas, provided that appropriate questions are asked, the right data needs are identified, and sensitive tools are used.

Placing the evaluation within the context of ILO's normative/policy frameworks for addressing discrimination

As a specialised agency of the United Nations, with the mandate for labour and social justice issues, the ILO has a normative function which is expressed through the adoption, ratification, supervision and implementation of International Labour Standards (ILS). Therefore it is a rights-based labour law-based organisation. This is the ILO's strongest asset and a comparative strength among all other international agencies. The elements of adoption, ratification, and supervision are integrated into a supervisory machinery. For the fourth element of implementation, the ILO also has technical departments, which provide expertise to constituents on best practices, with toolkits, research and expert advice, with the ILS as the major underlying structure.

The ILO has a tripartite structure unique in the UN in which employer and worker representatives, as the social partners of the economy, have an equal voice with those of governments in shaping its policies and programmes. The ILO encourages this tripartism within its member States by promoting social dialogue between trade unions and employers in formulating, and where appropriate, implementing national policy on social, economic and other issues.

The Constitution preamble clearly states ILO's role in the 'establishment of universal and lasting peace': ILO's goals are very much part of the Universal Declaration of Human Rights and inter alia various ILO covenants are contained in the UN Covenants on Civil and Political Rights and Economic, Social and Cultural Rights.

A benefit of the ILS is to provide clear rules to the global economy to ensure that economic progress will go hand in hand with social justice, prosperity and peace for all. Other than the UN system as a whole, no other development institution combines the standard-setting and development vocations.

Although the implementation of ILO's strategy to support member States in improving the impact of standards is shared across these many areas of the ILO, it is the International Labour Standards Department (NORMES) that anchors the strategy administratively and technically.

A. *Declaration on Fundamental Principles and Rights at Work (1998)*

A major milestone in the evolution of the ILO's commitment to non-discrimination was the ILO Declaration on Fundamental Principles and Rights at Work that was adopted by the International Labour Conference (ILC) in 1998. It pledges all members to respect, promote, and realize in good faith the principles and rights relating to freedom of association, and the effective recognition of the right to collective bargaining; the elimination of all forms of forced or compulsory labour; the effective abolition of child labour; and the elimination of discrimination in respect to employment and occupation.

The Discrimination (Employment and Occupation) Convention, 1958 (No. 111), and the Equal Remuneration Convention, 1951 (No. 100) have already been ratified by 169 and 168 Members respectively.[2] However, as affirmed by the 1998 Declaration, all ILO members have an obligation to respect, promote, and realize the fundamental

---

2   With regard to the progress made towards universal ratification, see GB.306/LILS/6

principles and rights concerning discrimination in employment and occupation. With regard to gender equality, the Workers with Family Responsibilities Convention, 1981 (No. 156), and the Maternity Protection Convention, 2000 (No. 183) are also considered to be key Conventions. The ILO Conventions regarding migrant workers and indigenous and tribal peoples, together with those addressing employment and vocational rehabilitation of people with disabilities, are crucial to addressing the discrimination faced by these groups.

B. *ILO Supervisory Mechanism*

International labour standards are backed by a supervisory system that is unique at the international level and that helps to ensure that countries implement the Conventions they ratify. The ILO regularly examines the application of standards in member states and points out areas where they could be better applied. If there are new problems in the application of standards, the ILO seeks to assist countries through social dialogue and technical assistance.

The ILO has developed various means of supervising the application of Conventions and Recommendations in law and practice following their adoption by the International Labour Conference and their ratification by States. There are two kinds of supervisory mechanism: (i) the regular system of supervision, and (ii) special procedures.

The regular system of supervision comprises the following two ILO bodies:

(i)   The Committee of Experts on the Application of Conventions and Recommendations (CEACR), which examines the application of the International Labour Standards (ILS) and makes observations and direct requests. Observations contain comments on fundamental questions raised by the application of a particular Convention by a state. Direct requests relate to more technical questions or requests for further information.

(ii) The annual report of the Committee of Experts, usually adopted in December, which is submitted to the International Labour Conference the following June, where it is examined by the Conference Committee on the Application of Standards (CCAS). The Conference Committee, a standing committee of the Conference, is made up of government, employer, and worker delegates. It examines the report in a tripartite setting and selects a number of observations from it for discussion. The

governments referenced in these comments are invited to respond before the Conference Committee and to provide information on the situation in question. In many cases, the Conference Committee draws up conclusions recommending that governments take specific steps to remedy a problem or to invite ILO missions or technical assistance.

Unlike the regular system of supervision, the special procedures mechanism comprises three special procedures based on the submission of a representation or a complaint: (i) Procedure for representations on the application of ratified Conventions, (ii) Procedure for complaints over the application of ratified Conventions, and (iii) Special procedure for complaints regarding freedom of association (Freedom of Association Committee).

Articles 24 and 26 to 34 of the ILO Constitution govern the complaint procedure. A complaint can be submitted, by another member State that has ratified the same Convention, against a member State for not complying with a ratified Convention. Upon receipt of a complaint, the Governing Body may form a Commission of Inquiry, consisting of three independent members. They are responsible for carrying out a full investigation, ascertaining all the facts of the case, and making recommendations on measures needed to address the problems raised by the complaint.

Improving the use of information emanating from the supervisory machinery

The reporting machinery puts considerable demands on governments, and comments are not always easy to interpret by those without a legal background. Governments reported that the recipient ministry of labour faced heavy demands for information that may require costly surveys. Governments also report duplications in the requests for information from the ILO.

Access to the wealth of information available provided by the supervisory bodies has improved through more accessible database and search functions. However, there is still scope for better and wider use of the information. Programme staff supporting the implementation, monitoring and reporting functions for non-discrimination need to be familiar with, and have some practical expertise on, how to support countries in applying Conventions. This is particularly important where specialists in International Labour Standards are not easily accessible due to competing demands from other coun-

try offices. Improved access to specialists in decent work teams, as a result of the field restructuring, will take time to implement.

Some good practices in translating the supervisory comments and observations into action include imaginative work by staff in the areas of gender and disabilities. This has empowered women and the disabled to push for their rights, as manifested through Convention No. 111 (Discrimination). In other areas, projects supporting the development of labour policies and improvement of labour relations have provided meaningful contributions to the ratification and implementation of relevant Conventions.

The evaluation mapped out the correlation between a sample of 82 technical cooperation operations approved and implemented between 2003 and 2011 and the comments from the ILO supervisory body. The analysis indicates a high degree of relevance (22 per cent mostly relevant and 65 per cent fully relevant) to the comments and observations of the supervisory bodies. Lessons from the field show, however, that generally it is difficult to mobilize extra-budgetary resources for normative activity.

Six core evaluative questions guided the analysis. These were:

(i)   To what extent is the ILO strategy relevant to the global and national policy dialogue addressing discrimination in employment and occupation?

(ii)   To what extent is the ILO strategy coherent and complementary, promoting synergies with other strategic outcomes, national constituents' priorities and partners to support Human Rights (HR) and Gender Equality (GE) in country programme objectives?

(iii)   To what extent does the ILO discrimination strategy lend itself to efficient implementation of HR and GE normative and policy frameworks?

(iv)   How effective is the strategy in addressing issues raised by ILO supervisory bodies and Global Reports (GRs) regarding non-discrimination in employment and occupation as a basic component of the HR and GE strategic and policy framework?

(v)   What impact did ILO actions have on policy, legal frameworks and awareness-raising regarding discrimination in employment and occupation as defined in HR and GE international and national normative and policy frameworks?

(vi)   To what extent were the ILO strategy and its means of action (technical cooperation projects and direct technical advice) designed and implemented to maximize sustainability of HR and GE gains at the country level?

The premise of the ILO evaluation was that employment discrimination is a violation of human rights. This premise establishes the basis for assessing the application of the ILO's principles and strategy regarding employment discrimination, and its implementation of those principles. The ILO evaluation team also reviewed specific areas of discrimination to reflect the categories of discrimination enumerated/encapsulated by international human rights law, focusing on gender equality, and discrimination based on race; sex; sexuality; religion; national origin; citizenship; and migrant status.

Twelve case studies illustrating the application of the ILO's strategy regarding discrimination in employment and occupation at the national level were performed in 11 countries together with one study on the Roma population in Europe,[3] in order to assess the application of relevant ILO Conventions. These were based on desk reviews of key country programmes, project reports and technical cooperation (TC) portfolios organized by country.

Criteria for the selection of case studies

A basic criterion for the selection of the country case studies was that, as far as possible, the countries selected are implementing a Decent Work Country Programme (DWCP) which includes projects or activities dealing with various aspects of discrimination and equality in employment. A second criterion considered the role that the ILO employment discrimination strategy played in promoting HR & GE principles of non-discrimination in the world of work. A third criterion considered the role of the ILO's elimination of discrimination strategy in the context of a particular DWCP. Finally, the selection of country case studies aimed for a balanced regional representation.

Based on these criteria, the evaluation team selected the case studies and categorized them into five thematic tracks:

- *Support for the promotion, implementation and application of the ILO Convention on Discrimination (No. 111)*

  Focus: Determine effectiveness of the interventions on the drafting of non-discrimination legislation and guidelines for their application and enforcement.

---

3   Case Studies: China, Jordan, Lebanon, Roma population in Europe, Philippines, Kenya, Moldova, Ukraine, Mali, Namibia, Benin, and Burkina Faso

- *Discrimination against indigenous peoples*

  Focus: Review ILO work in promotion and advocacy for the ratification of ILO Convention No. 169. Review the results and assess the effectiveness of ILO activities for promoting and protecting the cultural heritage of indigenous peoples in the world of work and occupation. Each of the 11 case studies reviewed actions taken to follow up the comments of the Committee of Experts on the Indigenous and Tribal Peoples Convention, 1989 (No. 169). The case study also assessed the effectiveness of the several training activities undertaken to strengthen capacities and develop advocates within government and the trade unions.

- *Gender discrimination and domestic workers and migration*

  Focus: Assess the extent to which the activities carried out to improve knowledge of selected ILO Conventions and recommendations relating to gender equality, namely Conventions Nos. 100, 111, 156 and 183, can be implemented in practice, in the context of each region, with a view to promoting ratification and improved implementation. In addition, identify trends and key responses to address the situation of domestic workers in the regions as it relates to the standard-setting discussion at the ILC.

  In the case of migration, the case studies sought to identify ILO support to tripartite efforts to discuss the comments of the Committee of Experts on Conventions Nos. 100, 111, 97 and 143 on equality of treatment for migrant workers, with a view to improving national legislation on discrimination in employment and occupation, specifically regarding migrant workers engaged in domestic work.

- *Discrimination based on HIV/AIDS status*

  Focus: Assessed the extent to which the ILO's programme on HIV/AIDS workplace education is achieving its intended objectives. Special attention was given to the participation of employers and workers organizations.

- *Discrimination based on disabilities*

  Focus: Assessed the effectiveness of ILO activities in promoting, through legislation, the employability of the people with disabilities. Identify lessons learned and good practices.

The amount of information reviewed/data collected was vast and varied as it covered a combination of primary and secondary data col-

lected through document reviews, interviews of stakeholders, online questionnaires, focal groups and triangulation. This approach was opted for given that budget and time constraints limited the number of field visits and respondents to surveys. The sampling strategy also included women and men in diverse stakeholder groups. Thus, the data reviewed came from multiple sources. This made it necessary to establish cross verification of information. For example, in the case of China, where rights-holders reported increased success in negotiating their needs or representing their interests, this was confirmed by reviewing records of decisions, or asking duty bearers if they have noticed any changes in the negotiation process with rights-holders. If HIV infected workers reported decreased stigmatization in the work place, this information was verified against records and reports produced by the Beijing YIRENPIN Centre, an NGO that advocates for the rights of HIV/AIDs and Hepatitis B infected persons.

The country case studies reviewed the relationship between the supervisory process and supporting member States in promoting the application of anti-discrimination laws and policies. Countries were chosen on the basis of geographic representation; type and complexity of non-discrimination issues at stake; and (c) the intensity of TC activities in the country.

In addition, field visits were conducted to one country, China, where interviews were held with ILO field staff, duty-bearers (Ministry of Labour officials and employer organizations) and rights-holders (trade unions, law advocacy organizations, and civil society organizations).

The case studies examined the ILO's work on gender discrimination; the promotion, implementation and application of Convention No. 111; and discrimination against domestic workers, migrant workers, indigenous peoples, people with HIV/AIDS, and the disabled. The Conventions most relevant to the studies are the Equal Remuneration Convention, 1951 (No. 100), the Discrimination (Employment and Occupation) Convention, 1958 (No. 111), the two Gender Equality Conventions (No. 156 and No. 183) and other Conventions that have an important impact on non-discrimination and equality (No. 97, No. 143, No. 159, and No. 169).

Involving stakeholders, implementers, duty-bearers and rights-holders: the case of China

In China, the ILO evaluation team interviewed officials of the Ministry of Human Resources and Social Security who were involved in the implementation of the 'Support to promote and apply ILO Convention No. 111 project'. A series of training programmes were con-

ducted under this project targeting policy- and law-makers, labour inspectors, employers and people involved in the settlement of employment discrimination disputes. A new web site and a newsletter were launched in 2009.

The evaluation team also interviewed ILO Country Office management and project managers; UN country team members; donor country representatives; service providers (i.e. All China Women's Federation; Employment Service and Administration Centre of China Disabled Persons' Federation; Beijing Zhicheng Migrant Workers' Legal Aid and Research Centre); Civil Society Organizations; and academics and activists working on issues of discrimination. The duty-bearers (including service providers) and rights-holders were selected based on the key thematic areas identified (gender discrimination; migrant discrimination; health status discrimination; sexual orientation and race/culture discrimination).

When interviewees were asked which groups were most susceptible to discrimination: 65.6 per cent responded that it was disabled people; 62.8 per cent responded that it was people with HIV/AIDS; 54.2 per cent responded that those infected with hepatitis B were most likely to face discrimination; 10.6 per cent responded that migrant workers were likely to be discriminated against in employment; and 6.3 per cent responded that discrimination based on sexual orientation was on the rise.

Interviewees were asked which of the following approaches were considered to be most effective to combat discrimination: government legislation; public policy and regulation; non-governmental organizations; mass media; or campaigns run by victims of discrimination. From the list, 65.2 per cent selected public policy and regulation. The second most popular option was government legislation (62.7 per cent) and the third was mass media (59.1 per cent). Only 22.8 per cent had confidence in campaigns launched by victims of discrimination.

Main findings on relevance, effectiveness, efficiency, coherence and sustainability

Relevance of the strategy to the problem it intends to address

The ILO strategy for eliminating discrimination in employment provides a global sense of direction for the implementation of non-discrimination work at the ILO. It is in line with the guidance on

non-discrimination provided through the follow-up mechanism for the ILO's 1998 Declaration on Fundamental Principles and Rights at Work and the 2008 Declaration on Social Justice for a Fair Globalization. It also addresses the issues relevant to discrimination highlighted in the Global Jobs Pact.

The evaluation noted that ILO technical cooperation action plans regarding the elimination of discrimination in employment, show a high degree of relevance to the discussion of the second Global Reports and the comments made by the ILO supervisory bodies. Most of the activities carried out under the Action Plan 2007–11 have supported the development of tools aimed at promoting non-discrimination in employment and occupation. However, the plan should focus more on new challenges arising from the global crisis.

Figure 1. TC Relevance to action plans
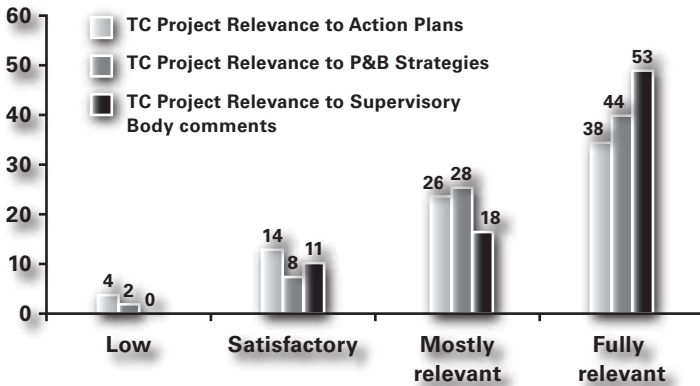
Note: TC projects relevance to Action Plans, P&B strategies and Supervisory Body comments (2003-2011, sample of 82 projects)

The case studies of this independent evaluation confirmed that the design of most non-discrimination activities was relevant to normative and policy frameworks for Human Rights & Gender Equality (HR & GE), as defined by the discrimination Conventions Nos. 100 and 111; national policies and strategies; and the needs of rights-holders and duty-bearers, both women and men, targeted by an intervention. Most results of the intervention were relevant to the realization of HR &GE. Some examples of areas to assess include the:

• extent to which the intervention is aligned and contributes to the implementation of international instruments such as the

Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), the Convention on the Rights of Persons with Disabilities (CRPD), the Convention on the Rights of the Child (CRC), and standards and principles on HR & GE;

• extent to which the intervention is aligned with and contributes to regional conventions and national policies and strategies on HR & GE;

• extent to which the intervention is informed by substantive and tailored human rights and gender analyses that identify underlying causes and barriers to HR & GE;

• extent to which the intervention is informed by needs and interests of diverse groups of stakeholders through in-depth consultation;

• relevance of stakeholder participation in the intervention.

The evaluation took advantage of existing data by making use of existing national or international data sets (on employment, income, vulnerability, disease, mortality, human rights violations, etc.) to compare and confirm or refute programme findings. This was particularly useful given the limited work that could be afforded. For example, the evaluation used results of gender and ethnic 'audits' conducted by the ILO in a number of countries. The goal of the audits was to ascertain whether, and how gender equality and the rights, needs and aspirations of indigenous and tribal peoples were taken into account, and if they were involved in the consultations leading to the formulation of PRSPs. These audits included Bangladesh; Bolivia; Cambodia; Guyana; Honduras; Kenya Lao PRD; Nepal; Nicaragua; Pakistan; Sri Lanka; Tanzania; Viet Nam and Zambia. The ethnic audit showed significant differences between regions and within regions, and between countries in terms of whether and how indigenous questions are addressed.

The ILO's work on non-discrimination has offered many opportunities for expanding its traditional network of national and international partners. In the case of China, it has been able to strengthen working relationships with other UN partners, line ministries and Civil Society Organizations (CSOs) through participation with the UN Joint Programme, 'China Youth Employment and Migration' (YEM). By participating in this joint programme with eight other UN agencies, the ILO has been able to establish strong alliances with the National Development and Reform Council (NDRC) and the Ministry of Civil Affairs (MOCA). They have also promoted research on

social inclusion of migrant workers and their families and are engaging stakeholders, especially the migrants themselves, with policy recommendations partly reflected in relevant national policies and plans. The CSOs have been an important part of the programmes and the ILO has established strong relationships with some of them. This has exposed CSOs to the ILO through engaging them in policy dialogue, training events, and other activities. These will positively impact on the evolution of stronger non-discrimination polices and laws.

The All China Women's Federation (ACWF) and Beijing University Women's Law Studies and Legal Aid Centre, with support from the United Nations Development Fund for Women (UNIFEM) and the United Nations Education, Scientific and Cultural Organization (UNESCO), surveyed 3,000 domestic workers to better understand their unique situation and needs. Recommendations have been made for subsidized vocational training; a job information system for migrant workers; the inclusion of migrants and their children in the urban social security system; and, a new law to govern domestic work and protect their labour rights.

Coherence of the strategy

The strategy's alignment with the recommendations of the second GR is explicit in the section on Experience and Lessons Learned. This section discusses the expansion of prohibited grounds of discrimination and an increased understanding of multiple discriminations. When addressing multiple discriminations, however, the strategy does not highlight emerging key linkages. For example, those seen among migration, ethnicity, social origin and gender, are addressed through the domestic workers programmes. Similarly, the emerging issues of age and gender, with respect to perceived inequities in pensions and retirement ages of women and men, are not highlighted. The strategy does not mention grounds for discrimination which receive inadequate attention, such as discrimination on the basis of race, mentioned in the first GR (Time for Equality at Work, 2003) as a priority issue.[4]

Guidance on how to establish synergies among approaches and tools developed in response to specific issues of discrimination is

---

4   Constituents at the ILC, 100th Session, Geneva, in June 2011 raised the issue of the scant attention paid to race discrimination. The results of the questionnaire survey confirm this: only 23 respondents (10 per cent) mentioned that discrimination on grounds of race was being addressed in their country of assignment, although discrimination on grounds of race may not be an issue in every country.

missing in the current strategy. This could optimize complementarities and synergies among the different products of ILO work in non-discrimination, thus conserving scarce resources.

The training package developed by the project on Convention No. 111 in China has been successfully applied to other projects for global use by the ILO NORMES and DECLARATION sections. However, the links between ILO work on the gender pay-gap and the treatment of other discrimination issues, presents an excellent opportunity for implementing lessons learned. The work on assessing the extent of discrimination experienced by migrant workers is good practice that could be transferred and adapted to other areas of discrimination work. These and other similar examples could be highlighted in the strategy for ILO's Strategic Outcome 17, to provide more guidance to field offices.

Regarding international partnerships, the strategy recognizes that:

> Non-discrimination is a key aspect of the human rights-based approach which is one of the common programming principles of the UN system. The Common Country Assessment and UNDAF documents, in an increasing number of countries, mainstream principles of non-discrimination and gender equality and pay particular attention to groups subject to discrimination and exclusion. The ILO will make particular efforts to align Decent Work Country Programmes with UNDAFs in this respect.[5]

Interviews with the ILO office in China and with their development partners, revealed that staff of all agencies felt their capacity was enhanced and enriched by working on joint programmes within the UNDAF, despite the usual caveats on transaction costs. These benefits of inter-agency collaboration were more broadly confirmed through the questionnaire survey.

Several respondents voiced concern about the potential for overlap between agency mandates. Training in gender analysis and mainstreaming was frequently cited along with the need for agencies to recognize a better division of labour in this field. The ILO has taken the lead in some aspects of gender mainstreaming within the UNDAF framework, with its Participatory Gender Audit (PGA) being requested by other UN agencies in a number of countries.[6]

---

5   Outcome 17 Discrimination, outcome-based work plans , page 12.

6   Gender Mainstreaming in DFID/ILO Partnership Agreement 2006-2009: Evaluation Report 16 November 2009 (GLO/08/53/UKM) Una Murray, Independent Consultant.

Effectiveness of the strategy

Effectiveness relates to how the strategy is implemented: how it brings together the multiple work-strands that contribute to the ILO's work on non-discrimination; how knowledge is shared; and, how well departments collaborate with each other. As support to the stated objective of the strategy, the evaluation assessed how the ILO functions and integrates its work-strands in enhancing the impact of non-discrimination activities. There are close connections between this and "focus and coherence". However, this relates more to actions taken in design, while "effectiveness" relates more to implementation and the attainment of desired results.

Over the last four years, several programmes have been developed and activities implemented at the global, regional and national levels. Non-discrimination has been included as a priority in the DWCPs of 36 countries. Particular attention has been paid to equal remuneration, the elimination of racial discrimination and better enforcement of legislation in general.

Promoting the rights of vulnerable groups, such as workers with HIV/AIDS or disabilities and indigenous peoples, are among the major areas of focus. Technical cooperation projects have included awareness raising; capacity building; information gathering and sharing; research; and training. The emphasis has been on providing advisory services and practical materials for capacity development.

With ILO support, more equality policies and action plans have been adopted and implemented at the national and workplace levels over the past four years. The work of bringing national legislation into line with the relevant ILO Conventions has continued. The ILO tools have been used more regularly; judges have increasingly referred to ILO Conventions in their case reviews; and, constituents are more aware of the perils of discrimination in the workplace and the overall economy.

The task team of the UNDAF Programming Network (UPN), led by ILO; the United Nations Children Fund (UNICEF); UNESCO; the Food and Agriculture Organization (FAO); the Office of the High Commissioner for Human Rights (OHCHR); and the UN Development Coordination Office (DOCO), recently conducted a desk review to determine the main areas covered in the UNDAFs, signed in 2010 for the period 2012–2016. Some of the main areas covered were human rights-based approach (HRBA), gender equality, capacity development and results-based management (RBM). The result of these show that in 15 of the 23 UNDAFs reviewed, the elimina-

tion of discrimination and achieving gender equality is a key objective. However, ILO involvement under these outcomes is explicit only in five of the 23.

The section in the strategy for Outcome 17 on Integration of HR & GE into decent work dimensions, confirms the ILO's interest in ensuring "that gender equality and non-discrimination[7] are addressed as cross-cutting issues in DWCPs under all strategic objectives". It also confirms that targeted action is taken to address the rights and needs of persons with disabilities, migrant workers, indigenous peoples, or persons living with HIV/AIDS.

The analysis of an intervention's effectiveness involved assessing the way in which results were defined, monitored and achieved (or not) on HR & GE and that the processes that led to these results were aligned with HR & GE principles (e.g. inclusion, non-discrimination, accountability, etc.).[8] In cases where HR & GE results were not explicitly stated in the planning document or results framework, assessing effectiveness in terms of HR & GE is still possible and necessary as all UN interventions will have some effect on HR & GE and should aim to contribute to their realization. Some issues to consider include the:

• extent to which the Theory of Change and results framework of the intervention integrated HR & GE;

• extent to which a human rights-based approach and a gender mainstreaming strategy were incorporated in the design and implementation of the intervention; and

• presence of key results on HR & GE.

Efficiency in the implementation of the strategy

The assessment of efficiency required a broader analysis of the benefits and related costs of integrating HR & GE in interventions. The evaluation team noted a number of issues with respect to efficiency. One main issue is training activities, as required in the strategy. The large number of training and capacity-building activities developed as products of non-discrimination interventions are not systematically recorded or disseminated for further use. The strategy for Outcome 17 does not address the utilization or dissemination of these products.

---

7   Programme and Budget for the biennium 2010-11, page 65.

8   Integrating Human Rights and Gender Equality in 26 Evaluation : Towards UNEG Guidance, page 26

It was frequently difficult to ascertain the number and categories of trainees who had been trained, how they had been selected, and how they would be applying the training subsequently.[9] The International Training Centre in Turin (ITC) has developed a database for its own training courses that could be developed and adopted elsewhere. A comprehensive database of training tools and materials was not easily accessible to headquarters or to field activities, even for individual issues of discrimination.

There were varying inputs from intended users for materials developed by headquarters and applied in the field. In the case of China, the Convention No. 111 manual developed for the country office, was extensively piloted in the country and was the result of inputs from constituents, the country office, the regional office and the ITC. In other cases, there was minimal input from the users, which undermined the efficient use of resources, and diminished the effectiveness of ILO interventions.

With respect to knowledge and knowledge management, the evaluation team noted the need for better management, developed under a multiplicity of individual and often small-scale research studies at headquarters and field locations. Research developed for a specific purpose (for example advocacy to the government to support employment of the disabled or to change the pension age for women, as under the ILO China programme) may be hard to retrieve and use when the specific purpose has been achieved and the programme ended. Nevertheless, such research should be considered an important part of the ILO's "situation analysis" for future work on non-discrimination, and could also contribute to the UNDAF database.

In recognition of the need to establish a more efficient approach to P&B work planning, and harmonize different parts of the ILO engaged in the same issues, but from different perspectives as called for in the Strategic Policy Framework (SPF), an outcome-based work plans (OBW) system was established in 2010. The coordination of this system relies on a network of Outcome Coordinators. The Coordinator for Outcome 17 is the Director of ILO DECLARATION section. The other units of the Office concerned with non-discrimination each have different reporting relationships.

---

9   Support to Promote and Apply ILO Convention no.111 Final Project Report, ILO Office for China and Mongolia, page 23.

Impact and sustainability

The long-term impact of the current strategy cannot be measured given its short implementation period. However, there are a number of significant immediate impacts at the country level. It is evident that the ILO's non-discrimination activities have made significant contributions to the efforts of national constituents to address discrimination in employment and occupation. Over the past four years, more equality policies and action plans have been adopted and implemented at the national and workplace levels due to ILO support. The work of bringing national laws into line with the relevant ILO Conventions has continued. The ILO tools have been used more regularly, judges have increasingly referred to ILO Conventions in their case reviews and constituents are more aware of their rights.

In order to help with the drafting of effective legislation, the ILO provides advice to governments and the social partners in the form of technical comments on proposed labour legislation. It also promotes good practice through its labour legislation guidelines.[10] Together with the ITC, the ILO delivers annual training on participatory labour law design and process, with particular attention to discrimination.

The sustainability of ILO non-discrimination work depends heavily on the ILO's ability to maintain its relevance through its advisory services, active research programme and capacity-building activities. This would help create strong institutions and effective mechanisms to ensure enforcement of ILS and fundamental rights and principles. The ILO's comparative advantage is not only its standard-setting capacity, but its capacity-building potential through close collaboration with the International Training Centre. This will help to continue to build the capacities of national constituents to address discrimination. It will do this by developing and applying training tools on labour inspection, gender equality and non-discrimination in the workplace through non-earmarked funds such as the Regular Budget Supplementary Account (RBSA).[11]

---

10   ILO: Labour Legislation Guidelines, last updated 10 Dec. 2001, available at http://www.ilo.org/public/english/dialogue/ifpdial/llg/index.htm, accessed 3 Feb. 2011.

11   Current ITC courses include: "Enhancing labour inspection effectiveness in selected countries in Europe and Central Asia" and "Strengthening labour inspection services in Angola, Brazil, China, India and South Africa". These training activities have also been offered at the national level, in Albania, Oman, Lebanon, The former Yugoslav Republic of Macedonia, and Yemen. In addition, guidelines concerning the role of labour inspection and the gender dimension in the workplace are being developed.

Conclusion

Evaluating the ILO's discrimination strategy through the human rights and gender equality lens has led to a careful assessment of the established protocol for conducting high-level strategy evaluations in the ILO. While all of ILO work is HR & GE based, the current protocol for evaluating its strategies weights heavily on the operational and organizational effectives in implementing its strategic strategies as presented in the biannual P&B document.

The discrimination strategy cuts across the work of all sectors of the organization, making the assessment of its impact a complex and daunting activity. The evaluation team relied on the information generated on achievement of outcomes as provided by comparing the ILO Implementation reports with target setting in the P&B and by interviews held with ILO staff to assess the effectiveness of the strategy.

The use of information from the P&B and Implementation Reports has strengths and weaknesses. On the plus side there is a historic value: a review of the information across and within ILO sectors and over several biennia gives an overall picture of how a large organisation such as the ILO with its many programme areas, is gradually changing its focus as it responds to Governing Body mandates, resource availability and changing approaches to meeting development challenges. Therefore, a review of this information has merit when assessing the performance of the ILO in mainstreaming such a central mandate as the ILO's strategy for eliminating discrimination.

But there are minuses. First, these are self-evaluation reports with indicator selection, target setting and outcome reporting, all within the hands of the implementing departments and the ILO programming machinery. There is no independent opinion as to whether events should move faster or be assessed on a more challenging basis. Second, there is the 'slice and dice' problem: that it is not easy, in two-year steps or slices, to representing the nuts and bolts of assessing progress on integrating the normative function in a specialised agency, which also aims to produce replicable good practices, tools and lessons for its constituents.

In order to offer diverse perspectives to the evaluation, and to promote participation of different groups of stakeholders, the HR & GE approach to the evaluation required setting-up an appropriate mix of qualitative and quantitative methods to gather and analyse data.

The evaluation used a mixed-method approach including (but not limited to) desk reviews, interviews, focus groups, surveys, etc. As described above. In order to ensure responsiveness to gender equality and human rights dimension in assessing the ILO's strategy to eliminate discrimination, the evaluation team took into account not only the policy and normative framework but also carefully discerned power relationships, and identified the structural causes of discrimination in employment and occupation.

Key message and lessons learned for future evaluations

The main message derived from the evaluation is that to tackle discrimination at work, the creation of more equal societies must become a central goal of development policies. The promotion of equal opportunities for decent work for all women and men, irrespective of race, religion, disability, age or sexual orientation, is one of the means to advance in this direction.

To evaluate the ultimate impact of the human rights-based strategies, would require a balanced mix of methods that would enable assessment of both organizational effectiveness in the implementation of the strategy and the relevance, impact and sustainability of the strategy's results at the national level.

References

Adib, A., and Guerrier, Y. The interlocking of gender with nationality, race, ethnicity and class: the narratives of women in hotel work. *Gender, Work, and Organisation*, Vol. 10, No. 4, pp.413-432

Bilge, S., and Roy, O. (2010). La discrimination intersectionnelle: la naissance et le développement d'un concept et les paradoxes de sa mise en application en droit antidiscriminatoire [Intersectional discrimination: The birth and development of a concept and the paradoxes of its implementation in anti-discrimination law]. In *Canadian Journal of Law and Society*, Vol. 25, No. 1, pp. 51-74.

Bowleg, L. (2008). When Black + Lesbian + Woman =/ Black Lesbian Woman: The methodological challenges of qualitative and quantitative intersectionality research. In *Sex Roles*, Vol. 59, No. 5.

ILO (2001). Labour Legislation Guidelines, last updated 10 Dec. 2001. Available at http://www.ilo.org/public/english/dialogue/ifpdial/llg/index.htm,

ILO (2003). Time for equality at work. Global rReport under the follow-up to the ILO Declaration on Fundamental Principles and Rights at Work.

ILO (2007). equality at work: Tackling the challenges. Global Report of the Director-General on the follow-up to the ILO Declaration on Fundamental Principles and Rights at Work. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---ilo_aids/documents/publication/wcms_127385.pdf

ILO (2010). Programme and Budget 2010-11. http://www.ilo.org/public/english/bureau/program/download/pdf/10-11/pb.pdf

ILO (2010). *Employers' organisations responding to the impact of the crisis.* Working Paper No.2. Bureau for Employers' Activites, Geneva. Available at:http://www.ilo.org/public/libdoc/ilo/2010/110B09_17_engl.pdf

ILO (2010). Recommendation concerning HIV and AIDS and the World of Work, (No. 200). Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---ilo_aids/documents/normativeinstrument/wcms_142706.pdf

ILO (2011). Independent Evaluation of the ILO Strategy for the Elimination of Discrimination in Employment and Occupation. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165234.pdf

ILO (2011). Independent evaluation of the ILO's strategy to address HIV and AIDS and the world of work. Available at: http://www.ilo.org/eval/Evaluationreports/Strategyandpolicyevaluations/WCMS_165849/lang--en/index.htm

ILO Fundamental Principles and Rights at Work. Discrimination at Work in Asia. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_decl_fs_89_en.pdf

ILO Fundamental Principles and Rights at Work. Discrimination at Work in Africa. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_decl_fs_88_en.pdf

ILO Fundamental Principles and Rights at Work. Discrimination at Work in Africa. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_decl_fs_90_en.pdf

ILO Recovering from the crisis: A Global Jobs Pact. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---relconf/documents/meetingdocument/wcms_115076.pdf

ILO, Declaration on Social Justice for a Fair Globalization. Available at: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_099766.pdf

ILO Joint Statement by the Business Organisations of the G-20 Countries (B-20) and labour organisations of the G-20 Countries (L-20). Available at: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/meetingdocument/wcms_166713.pdf

ILOSupport to Promote and Apply ILO Convention no.111 Final Project Report, ILO Office for China and Mongolia.

ILO Constitution, ILO Database of International Labour Standards (ILEX). Accessed January 2012 at: http://www.ilo.org/ilolex/english/index.htm

ILO Convention Discrimination (Employment and Occupation) Convention, 1958 (No. 111). Accesses January 2012 at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_decl_fs_85_en.pdf

Miendjiem, I. (2011). Etude nationale sur la discrimination en matière d'emploi et de profession et proposition d'un plan national d'action au Cameroun, ILO.

Sheppard, C. (2011). Multiple discrimination in the world of work. International Labour Office, Programme for the Promotion of the ILO Declaration on Fundamental Principles and Rights at Work (DECLARATION), Geneva.

CONEVAL EXPERIENCE IN EVALUATING INTERVENTIONS FOR INDIGENOUS POPULATION

*Gonzalo Hernández Licona, Thania de la Garza,
María Fernanda Paredes and Brenda Valdez,
Mexico's National council for the evaluation of social policies (CONEVAL)*

Mexico is a country with prevailing challenges in various dimensions with regard to social disparities. One of the harshest manifestations of the social gaps that persist in Mexico is the lack of opportunities for the indigenous population, which leads to serious limitations for the exercise of their rights and provides evidence of the social inequalities that prevail among the population. The second article of the Mexican Constitution sets forth the obligation of the State to guarantee the equality of opportunities for the indigenous population by improving their living conditions. However, a constant struggle remains when it comes to assess improvements in their living standards.

In order to analyze the situation of the indigenous people and the government response to deal with the problems that come up, we present an assessment of the adequacy and results of social programmes based on information provided by three tools generated by the National council for the evaluation of social policies (CONEVAL). These are the multidimensional poverty measurement; CONEVAL's inventory of social programmes; and the evaluation of programmes that target this population.

The *multidimensional poverty measurement* is based on articles 36 and 37 of the General Social Development Law, which stipulate that the guidelines and criteria issued by CONEVAL for the definition, identification and measurement of poverty, are mandatory to all federal states and institutions that participate in social development. Furthermore, the law establishes that CONEVAL should use the information generated by the National Institute of Statistics and Geography[1] (INEGI) for the calculation of, at least, the following eight indicators:

• Current per capita income;

• Average educational gap in the household;

---

1   Instituto Nacional de Estadística y Geografía, INEGI.

- Access to health services;

- Access to social security;

- Quality and spaces of the dwelling;

- Access to basic services in the dwelling;

- Access to food; and

- Degree of social cohesion.

Having taken into account the normative background, CONEVAL decided to consider separately the space of economic wellbeing and the space of social rights since their theoretical and conceptual nature differ. This led to the definition of a bi-dimensional measurement to identify the poor in which the first dimension relates to economic wellbeing and the second to deprivation of economic, social, cultural and environmental rights. According to CONEVAL, a person is in multidimensional poverty if he/she is deprived in both the economic wellbeing space and the rights space.

The importance of this measurement lies in the provision of detailed information that enables the identification of the target population of different social programmes. Before this measurement was developed, programmes used to be evaluated on the sole basis of income poverty, even when their actions had a distant relation with it[2]. Since 2009 when the methodology was defined, CONEVAL has measured poverty at national and state level and for different population groups for two periods: 2008 and 2010. Based on these poverty estimates, we have analyzed the social context of the indigenous group.

---

2   On the one hand, the human rights approach to poverty measurement is based on the recognition of human rights as "the expression of the needs, values, interests and goods that, because of their urgency and importance, have been considered fundamental and common to all human beings" (Kurczyn & Gutierrez, 2009: 3-4). Accordingly, everyone should have access to social rights that guarantee human dignity, which have been adopted within the national legal framework and ratified through the signing of international agreements that protect them. The State must create the mechanisms that will progressively allow full access to human rights. If any (some) of these rights are not satisfied, the person is considered to be deprived of this (these) rights. On the other hand, the fundamental objective of the wellbeing approach is to identify conditions that limit people's freedom to develop fully. This assumes that every person, according to his/her circumstances and preferences, develops a set of capabilities that define the range of life options he/she may choose from. If these options do not allow him/her to have acceptable living conditions within his/her society, the individual is considered to be in a deprived state of wellbeing.

The second tool used for this assessment is *CONEVAL's inventory of social programmes*, a database that integrates information – in 111 variables – from 273 federal social development programmes. The information has been associated[3] with the access to social, cultural and environmental rights established in the General Social Development Law, such as education; food; labour; equality; housing; health; economic wealth; social security; and wholesome environment. Therefore, one of the main strengths of this database is that it enables the recognition of the set of interventions, which each contributes on a different scale, to social development.

In this context, CONEVAL's inventory is useful to identify programmes that share similar information, by using a definition of social programme and criteria for including or excluding programmes in the database[4]. In this paper we use CONEVAL's inventory to analyze the group of programmes that focus on the indigenous people.

According to Mexico's institutional evaluation frameworks, various types of tools can be used to evaluate social programmes. One of them is the *Specific Performance Evaluation* (EED, by its Spanish acronym), a nine-page report in which the assessment of the annual performance of social programmes is summarized. This evaluation describes the most relevant features of programmes performance, such as achievement of outcomes, outputs, coverage and follow-up of evaluations recommendations, according to the expert opinion of an external evaluator. In 2010, CONEVAL coordinated 133 EEDs of social programmes; among which nine were targeted on the indigenous population.

CONEVAL carried out a strategic assessment, the *Comprehensive Performance Evaluation,* for each of the different arenas analyzed by the EEDs with the purpose of reaching a holistic analysis. This evaluation depicts the overall relevance of programmes that address common problems, as well as its complimentary, strengths and opportunity areas, in order to improve coordination between them. One of the *Comprehensive Performance Evaluations* developed by CONEVAL in 2010 analyzed nine programmes focused on the indigenous population. We used this assessment, along with the EEDs,

---

3   The Inventory states two types of links: the direct one - associates directly the alignment between the programme´s main objective and the social right, and the indirect one - associates the relative degree of causality between the main objective and the social right.

4   The definition of federal social programme and the criteria for the inclusion of programmes in CONEVAL's inventory can be consulted at http://web.coneval.gob.mx/Informes/ArchivosSIPF/Nota_Metodologica.pdf

to examine the efficiency of government responses to alleviate problems that affect the indigenous population.

Poverty in the indigenous population (2008 – 2010)

According to *INEGI*, six per cent of Mexico´s total population, almost 7 million people, was indigenous in 2010.[5] People from this population were found in 28 out of the 32 Mexican states; however, the majority was located in the states of Oaxaca, Chiapas, Veracruz, Yucatan, Puebla and Mexico.
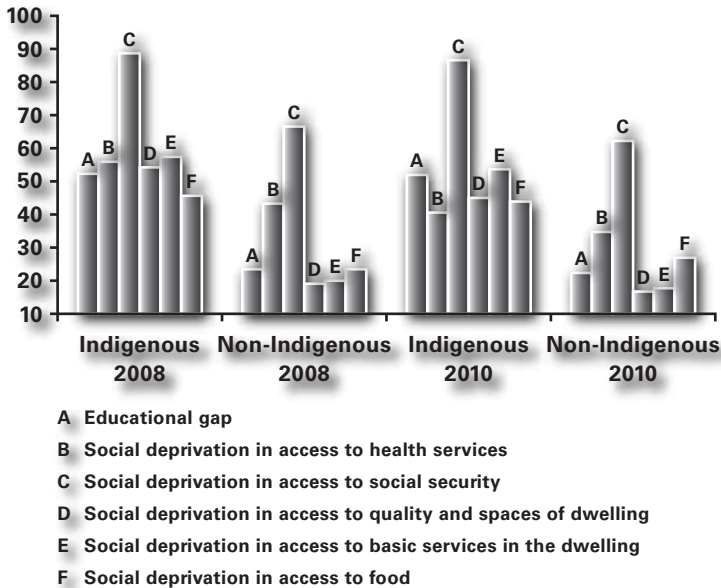
Moreover, according to CONEVAL estimates, the indigenous population living in poverty increased from 5.3 million in 2008 to 5.4 million in 2010. Consequently, in 2010, 79.3 per cent of the indigenous population was poor, among which 40 per cent was extremely poor and the remaining 39 per cent, moderately poor[6].

Graph 1 shows the percentage of deprived people in the indigenous and the non-indigenous population in each of the indicators considered in Mexico´s poverty measurement. The social deprivation indicators bars show that the indigenous population experienced less deprivation in 2010 than in 2008. However, the differences between the indigenous and the non-indigenous population are still considerable. For instance, 94 per cent of the indigenous population suffers from at least one social deprivation and 64 per cent live with at least three social deprivations simultaneously. In contrast, only 24 per cent of the non-indigenous population lives with more than three social deprivations.

---

5   According to INEGI, a person is indigenous if he/she speaks one of the 68 native languages that the Instituto Nacional de Lenguas Indígenas of Mexico recognizes http://www.inali.gob.mx/clin-inali/.

6   CONEVAL published the methodology for measuring multidimensional poverty in Mexico in December 2009 and to perform the calculations uses information from the Socioeconomic Conditions Module 2010 of the National Household Income and Expenditures Survey generated by the National Statistics Institute.

Graph 1: Social Deprivation Indicators for Indigenous and Non Indigenous Population, 2008-2010

A Educational gap

B Social deprivation in access to health services

C Social deprivation in access to social security

D Social deprivation in access to quality and spaces of dwelling

E Social deprivation in access to basic services in the dwelling

F Social deprivation in access to food

Furthermore, social disparities in income indicators present the same tendency as deprivation indicators. In 2010, while 79 per cent of the indigenous population had an income below the wellbeing threshold[7], this was so for only 50 per cent of the non-indigenous population. This last scenario does not change if the minimum well-being threshold[8] is considered, in which case 50 per cent of the total indigenous population reported having an income below this threshold. Moreover, from 2008 to 2010 both population groups experienced an increase in the percentage of people with income below this threshold; however, the increase was 4.5 percentage points higher for the indigenous group.

Various social indicators evince prevailing inequalities among these groups. In the case of education, 27.3 per cent of the 15 years or older indigenous population is illiterate, against 10 per cent of the

---

7   According to Mexico´s Methodology for Poverty Measurement, this is the population that does not have sufficient resources to acquire the necessary goods and services to satisfy their needs (food and non-food).

8   According to Mexico´s Methodology for Poverty Measurement, this is the population that, even when using all of their income to purchase food, cannot acquire enough of it to ensure adequate nutrition.

non-indigenous population; almost one out of every five indigenous persons do not have access to education, among which 13 per cent of the population between 6 and 14 years do not attend school, and less than 3 per cent of the youth have access to undergraduate education. Likewise, the number of the indigenous population with social deprivations is very high: 83.5 per cent of the indigenous population lacks access to social security, 50.6 per cent to basic services in the dwelling, 42 per cent to quality and space of the dwelling, and 40.5 per cent to food. What is more, malnutrition in indigenous children is twice as high as the national average. Finally, regarding the economic wellbeing dimension, the gap in income between the indigenous and the non-indigenous population varies around by 36 per cent. The employment rate for the indigenous population rises to 48 per cent, and, within it, 23 per cent do not receive income and 53.5 per cent receive an income lower than the minimum wage.

Comprehensive Performance Evaluation of the Government response to the socioeconomic situation of the indigenous population in Mexico

According to CONEVAL's inventory of social programmes, in 2010 there were seventeen federal social development programmes aimed at reducing the inequality gap of the indigenous population. These programmes were operated by the National Commission for the Development of the Indigenous Population (CDI), the Ministry of Public Education (SEP), the Ministry of Social Development (SEDESOL) and the Mexican Social Security Institute (IMSS).

Each programme provides different types of supports to the indigenous population and is associated with different social rights, as it is shown in Table 1. The majority of programmes are linked (directly and indirectly) to the social right of non-discrimination, followed by the right of employment, education and health as can be seen in Graph 2.

Graph 2: Social rights addressed by the Programmes
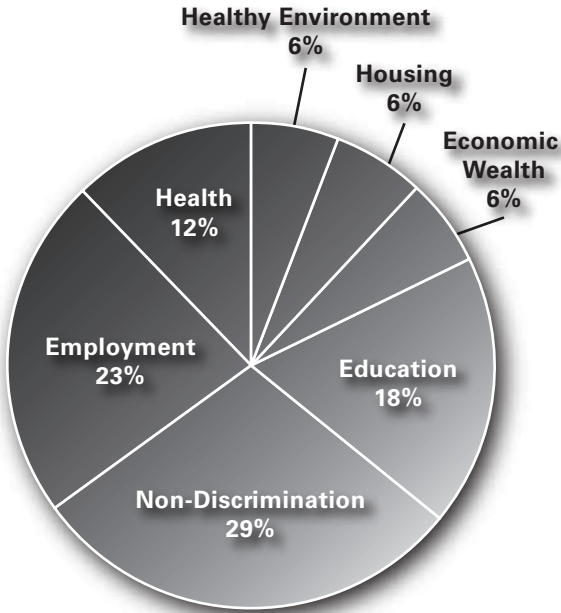
This is consistent both with the government goals of improving equality, social justice and economic wellbeing, and with the objective of increasing development, since the United Nations Development Programme states that the most important variables for increasing development are per capita income, education and life expectancy.

Graph 3 presents the 2010 programme budgets associated with social rights.
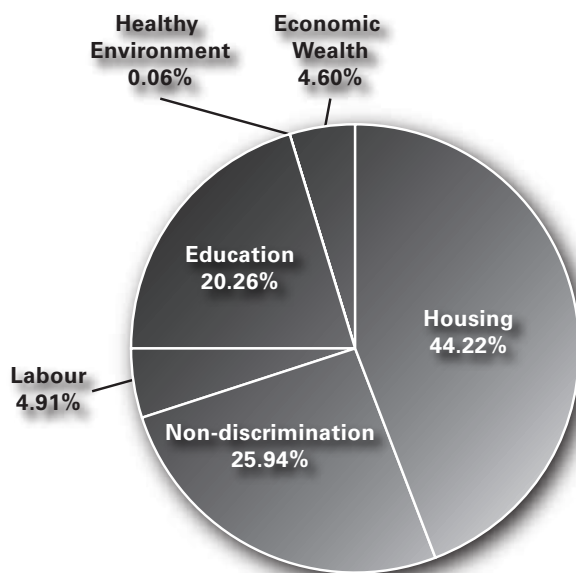
Social Right	Executed Budget 2010
Non-discrimination	$ 4,275.86
Labour	$ 809.07
Education	$ 3,339.33
Healthy Environment	$ 9.93
Economic Wealth	$ 758.10
Housing	$ 7,288.56

Source: CONEVAL's Inventory of Federal Programmes and Actions of Social Development 2010.
Millions of pesos 2010.

In this section, the analysis is based on the programmes that are included in the Comprehensive Performance Evaluation. Together, the programmes address issues related to educational services, cultural development, access to justice, and attention to displaced groups. Four of them (*PAEI; Programa de Educación Inicial y Básica para la Población Rural e Indígena; PRONIM; and PATP*) are targeted at educational goals, such as diminishing the inequality of access, increasing the stay in school, and completing education cycles. Each programme has its own strategies; nevertheless, evidence concerning the coordination between programmes and linking public policies among them is insufficient.

The rules of operation and the policies of the programmes do not take into consideration the indigenous urban population, which represents almost 50 per cent of the indigenous population, nor the intercultural indicators to measure the progress of their beneficiaries. Additionally, the population is hardly involved in their design, implementation and monitoring. However, there is no evidence of counterproductive effects on the programmes' beneficiaries, despite the lack of coordination among them.

In terms of access to justice, the objective of the nine programmes is to create the necessary conditions for the indigenous population to exercise their rights, both as political groups and individually. The programme *Promoción de Convenios en Materia de Justica (PPCMJ)* is a key element to achieve this objective since the lack of trained indigenous language translators and defenders causes the spread of discriminatory practices and corruption, among others things. In this regard, the CDI performs other actions that enhance the programme results, such as the release of prisoners and training indigenous legal translator.

According to the EEDs, the programmes do not change the inequality of opportunities faced by the indigenous population; instead, they implement strategies and actions to start development processes (that do not match the actions undertaken by other public dependencies), by boosting the dialogue with a culturally diverse population that has been socially excluded and that is represented by organizations with different management capabilities. The necessity and relevance of the nine programmes that belong to this theme group relies on the evidence that indicates that the poverty gap for the indigenous population is widening.

The programmes operate in isolation having little or no connection with their respective fields. This generates a proliferation of pro-

cedures and requirements for the target population that impedes access, especially for indigenous women. Thus, without a comprehensive indigenous policy, the actions of the nine programmes that belong to the group only partially contribute to reducing inequality of the indigenous population. There is no doubt about the relevance of these programmes; however, it would be worthwhile initiating actions that address these issues and to continue the efforts to measure the outcomes regarding the problems they are addressing.

Despite the efforts by the government to widen the eligibility criteria of the programmes and to target their actions for the indigenous population, the socioeconomic disparities concerned with this population group still prevail in several areas, such as access to basic social services; lack of main roads; social, economic and institutional lag in their places of residence; educational backwardness; and low economic resources, among others. Under these circumstances, the indigenous population does not have the necessary conditions to exercise their basic social rights and therefore to take full advantage of the benefits given by the federal social development programmes. For instance, it is clear that the current situation of the indigenous population has represented, and still represents, a difficult challenge for the Mexican government that requires changes or readjustments in several programmes, interventions, actions, projects, processes and strategies. Access to basic social rights that enhance the living conditions of the population group should be guaranteed. Thus, the development of sound programmes that can reduce and eliminate – in the long run - the socioeconomic gaps between the indigenous and the non indigenous populations have never been more urgent.

Finally, there are several findings, achievements and key challenges that can be drawn from the Comprehensive Performance Evaluations, carried out in 2010, for the programmes focused on the indigenous population:

Achievements

- The EEDs and the Comprehensive Performance Evaluations of the programmes targeted at the indigenous population have assessed their performance to explore their strengths and weaknesses. Likewise, they have demonstrated their relevance and validity in dealing with the issues that hinder the development of the indigenous population.

- The first overall success of the programmes is the continuity of their purpose because the available data demonstrates the persistence of the problems they are trying to alleviate.

- The CDI´s programmes encourage social community participation and revival of their traditions so that beneficiaries gain acceptance and recognition among the population.

Key Challenges

- None of the agencies has an integral planning framework to work together and link targets and coordinate their programmes (CDI and SEP), not only in terms of coverage and resources, but specially for continuity and complete attention. This lack of coordination affects the allocation of fiscal resources and their use as instruments for equitable development and cultural diversity.

- Given the magnitude of the indigenous development problems it is difficult to assume that the dependencies in charge of the programmes could meet the needs of the indigenous people; hence, the limited coverage. However, this cannot be considered as an institutional failure, but as the opportunity to increase the attention to this sector.

- The guidelines and rules of operation of the programmes do not cover the urban indigenous population, which represents an important influence in the demand for resources.

- The achievement of the programmes objectives requires intensive and specialized institutional work to enhance inclusion of the indigenous population in the institutional definitions and in the monitoring of the results.

- No impact evaluation was carried out between 2008 and 2010 for these programmes; thus, the available information is limited. Programmes should produce feasibility studies to find out if it is possible to analyze and follow-up the impact on their beneficiaries.

Findings

- The 2010 EEDs showed that most of the programmes followed-up the recommendations from evaluations. Moreover, those with external evaluations initiated actions drawn from the recommendations. The three institutions evaluated demonstrated their ability to incorporate the evaluation results by improving their design, operation, and processes.

- According to the EEDs, programmes belonging to this theme group require more rigorous methodologies to measure the perception of the beneficiaries, the fulfillment of their goals, and, above all, the achievement of their objectives.

Table 1: Federal Programmes and Actions of Social Development aimed to the Indigenous population that makes up the CONEVAL Inventory.

Agency	Programme	Support Type	Direct Associated Social Right
	Programas Albergues Escolares Indígenas (PAEI)	Shelter/Other : school supplies, extracurricular activities, support for homework	Education
	Programa de Infraestructura Básica para la Atención de los Pueblos Indígenas (PIBAI)	Public Works	Non-discrimination
	Programa Fondos Regionales Indígenas (PFRI)	Legal advice/Training/Other : funding of productive projects or microcredit	Labour
	Programa Organización Productiva para Mujeres Indígenas (POPMI)	Technical Assessment/Scholarship/Training/Other: funding of productive projects or microcredit, Input for production, machinery equipment and tools	Labour
	Programa Promoción de Convenios en Materia de Justicia (PPCMJ)	Funding of social projetcs	Non-discrimination
CDI	Programa de Fomento y Desarrollo de las Culturas Indígenas (PROFODECI)	Funding of social projects	Non-discrimination
	Programa Turismo Alternativo en Zonas Indígenas (PTAZI)	Other : funding of productive projects or microcredit	Labour
	Programa de Coordinación para el Apoyo a la Producción Indígena (PROCAPI)	Technical Assessment/Training/Other : funding of productive projects or microcredit	Labour
	Proyecto para la Atención a Indígenas Desplazados Indígenas urbanos y migrantes (PAID)	Other : Crop lands, urban solars, construction materials for housing and, inputs for production.	Non-discrimination
	Apoyo a Proyectos de Comunicación Indígena (APCI)	Other : funding of productive projects or microcredit	Non-discrimination
	Atención a Tercer Nivel	Shelter/Food/Health Services/Other : Medicines	Health
	Manejo y Conservación de Recursos Naturales en Zonas Indígenas (MANCON)	Other : Economic resources to support specific projects in matter of natural resources conservation	Healthy environment

Agency	Programme	Support Type	Direct Associated Social Right
SEP	Programa Asesor Técnico Pedagógico (PATP)	Technical Assessment/Training/Other : Technology transfer to producers and students as well as equipments to schools related to rural production	Education
IMSS	Programa IMSS-Oportunidades	Health Services	Health
SEDESOL	Programa de Opciones Productivas (POP)	Other : funding of productive projects and development of capacities for production	Economic Wealth
	Programa para el Desarrollo de Zonas Prioritarias (PDZP)	New housing or improvement of the existent one/ Other : Projects of social basic infrastructure,	Housing
SEP	Programa de Educación Inicial y Básica para la población rural e indígena	Scholarships/Books/Education Services	Education

Source: CONEVAL Inventory of Social Development Programmes 2010.

256

References

Comisión Nacional para el Desarrollo de los Pueblos Indígenas (2009). 'Acciones de Gobierno para el Desarrollo Integral de los Pueblos Indígenas. Informe 2008', México.

CONEVAL (2009). Secretaria de Hacienda y Crédito Público y Secretaria de la Función Pública, 'Lineamientos para la Evaluación de los Programas Federales de Desarrollo Social', México, 2007. Available at: http://www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/normatividad/lineamientosgenerales.es.do;jsessionid=e1a48043b850289ddd76d0562114a5c27c2c2a3357af6cea7395a54d233ed603.e34QaN4LaxeOa40Nbx10

CONEVAL (2011). 'Evaluación Integral del Desempeño de los Programas Federales y Acciones dirigidas a la Población Indígena 2009-2010', México, 2011.

CONEVAL (2011). 'Evaluación Integral del Desempeño de los Programas Federales y Acciones dirigidas a la Población Indígena 2010-2011', México, 2011.

CONEVAL, (2011).'Informe de seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2010', México, 2011. Available at: http://www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/seguimiento_recomendaciones/ASM_2010_2011.es.do;jsessionid=e1a48043b850289ddd76d0562114a5c27c2c2a3357af6cea7395a54d233ed603.e34QaN4LaxeOa40Nbx10

CONEVAL (2011). 'Inventario CONEVAL de Programas y Acciones Federales de Desarrollo Social 2010', México.

CONEVAL (2011). 'Medición de la Pobreza 2010', Anexo Estadístico, Cuadro 13, México.

H. Congreso de la Unión Cámara de Diputados, Ley General de Desarrollo Social, México, 2004. Available at: http://www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/normatividad/leygeneral.es.do;jsessionid=e1a48043b850289ddd76d0562114a5c27c2c2a3357af6cea7395a54d233ed603.e34QaN4LaxeOa40Nbx10

Plan Nacional de Desarrollo 2007-2012, México (2007). Available at: http://pnd.calderon.presidencia.gob.mx/index.php?page=documentos-pdf

UNICEF-SUPPORTED EVALUATIONS WITH ELEMENTS OF EQUITY-FOCUSED EVALUATIONS

The following examples illustrate different ways in which equity-focused evaluations have been designed and used by UNICEF and its partners.

Evaluation of the UNICEF Education Programme in Timor-L'Este 2003-2009. "From Emergency Responses to Sustainable Development for Children and Adolescents in Timor-L'Este". Available at: http://www.unicef.org/evaldatabase/index_58819.html

> This case illustrates how equity issues can be addressed in a context where there is only limited access to quantitative data, and the evaluation must mainly rely on a mixed-method approach.

Evaluating the equity-outcomes of the Nepal Education for All Project. Available at: http://www.unicef.org/evaldatabase/index_58884.html

> The evaluation did not have a specific equity focus but national partners requested that the sample selection be target at some of the poorest and most remote communities, where ethnic minorities and other vulnerable groups represented a high proportion of the population.

Evaluating the equity outcomes of the Cambodia Community-Led Total Sanitation Project. Available at: http://www.unicef.org/evaldatabase/index_57963.html

> One of the central objectives of the project was to develop methodologies to ensure the participation of all sectors of the population, including the poorest and most vulnerable. A central goal of the evaluation was to assess the equity outcomes of the project.

Evaluating the impact of social assistance on reducing child poverty and child social exclusion in Albania. Available at: http://www.unicef.org/evaldatabase/index_59597.html

> This case illustrates how national data sets can be analyzed to prepare a typology of vulnerable groups who are not adequately supported by the national social safety net.

Inter-Agency Real-Time Evaluation of the Humanitarian Response to Pakistan's 2009 Displacement Crisis. Available at: http://www.unicef.org/evaldatabase/index_59598.html

This case illustrates how equity issues were addressed in the evaluation of the response by the international community to the humanitarian crisis created by a massive population displacement in Pakistan. It describes the use of a mixed-method approach that sought to ensure the credibility of the evaluation findings through the presentation of an evidence table and the systematic use of triangulation. It also documents the many political, security and logistical challenges in conducting an evaluation in a military emergency situation. The case illustrates the importance of an equity-focus as programmes were mainly planned in consultation with village elders and male household heads and little attention was given to the special needs of women and children and the poorest and most vulnerable families.

Evaluation of the Egyptian Community Schools Project. Available at: http://www.unicef.org/evaldatabase/index_59600.html

This case describes an Equity-focused evaluation that was specifically designed to assess the effectiveness of community-based schools in increasing school enrolment and performance for under-served population groups, with particular attention to girls. It also discusses the practical challenges of identifying a well-matched comparison group. Both quantitative and qualitative data collection methods are used but there is no discussion of how these are integrated into a mixed-method strategy or how triangulation is used to strengthen validity of the data, findings and conclusions.

Evaluation of the Tanzania Community Justice Facilitation Project. Available at: http://www.unicef.org/evaldatabase/index_59601.html

This case describes an Equity-focused evaluation that assesses the effectiveness of the community justice facilitation project in ensuring that justice is accessible to women and children. It combines quantitative and qualitative data collection methods but does not describe an integrated mixed-method approach or the use of triangulation to strengthen the validity of the data and findings. The practical challenges in conducting a rigorous evaluation design within a multi-level administrative system are also described.

**Evaluating UNICEF's Response in the area of Child Protection in
Indonesia, to the 2004 Indian Ocean Tsunami (2005-2008).** Available at: http://www.unicef.org/evaldatabase/index_59604.html

The evaluation, which was commissioned by UNICEF's Child
Protection Department, was aimed at determining the impact
of the UNICEF response to the tsunami within the child protection sector, and drawing lessons learned and recommendations
for both the recovery/transition and on-going development programming, and policies to improve the well-being and rights of
children and women. It follows the evolution of the three child
protection work strands (children without family care, psychosocial support, and exploitation and abuse) through the different phases of their development and it examines the extent to
which child protection results were achieved in each phase and
to which they are likely to be sustained.

Six cross-cutting issues were examined: a) advocacy, policy and
coordination; b) reaching the most vulnerable; c) gender; d) conflict; e) emergency, recovery, and early development linkages;
and f) child protection systems capacity development.

The evaluation employed a sequential mixed-methods approach
to combine comprehensive coverage with in-depth analysis.
It focused on three districts to enable comparison of results
between tsunami and conflict (mainly) affected districts, which
allowed for comparisons between those areas with a strong
operational UNICEF presence and those areas with less. The
evaluation design also compared different interventions with
one another – or, where a similar programme did not exist, with
groups of children who did not receive the intervention.

**Long-term evaluation of the Tostan programme to reduce
female circumcision in villages in three regions of Senegal.**
Available at: http://www.unicef.org/evaldatabase/index_59605.
html

The goal of the Tostan (a Senegalese NGO) programme was to
reduce the prevalence rate of female circumcision, to increase
age at first marriage and to improve the health status of mothers in villages in three regions of Senegal, through promoting
social change based on capacity building and participatory development. The long-term evaluation used a mixed-method design:
combining a quantitative district household survey covering
knowledge of female circumcision and prevalence rates, and

age at marriage and health status, with qualitative techniques to assess the programme implementation process, to understand how villages organized their participation in public declarations, and to obtain women's opinions about the impact of the programme. Three groups of villages were compared: villages that had benefited from a Tostan programme and had publicly declared that they would abandon the practice of circumcision; villages that that had made a public declaration to abandon female circumcision but did not benefit directly from a Tostan programme; and, a control group of villages that practice circumcision but had not been exposed to the Tostan programme.

Annexes

AUTHOR VITAE

BAMBERGER, Michael J. is an Independent Consultant in the United States. Michael Bamberger has a Ph.D. in sociology from the London School of Economics. He has 45 years of experience in development evaluation, including a decade working with NGOs in Latin America on housing and urban development, and almost 25 years working on evaluation and gender and development issues with the World Bank in most of the social and economic sectors in Latin America, Africa, Asia and the Middle-East. He has also had a decade as an independent evaluation consultant working with 10 different UN agencies, the World Bank and the regional development banks, bilateral agencies, developing country governments and NGOs. He has published three books on development evaluation, numerous handbooks on evaluation methodology, and articles in leading evaluation journals. He has been active for 20 years with the American Evaluation Association, serving on the Board and as chair of the International Committee. He has served on the Editorial Advisory Boards of the American Journal of Evaluation, New Directions for Evaluation, the Journal for Mixed Method Research and the Journal of Development Effectiveness, and is a regular reviewer for numerous professional evaluation journals. He is on the faculty of the International Programme for Development Evaluation Training (IPDET) where he lectures on conducting evaluations under budget, time and data constraints; the gender dimensions of impact evaluation; mixed method evaluation and the evaluation of complex programmes. He also teaches at the Foundation for Advanced Studies for International Development in Tokyo.

BLEDSOE, Katrina L. is a research scientist and senior evaluator at the Education Development Center, Inc., located in Washington D.C. She received her doctorate in psychology with a focus on applied social psychology, and a specialization in programme evaluation from Claremont Graduate University. Her expertise is in applied social psychology, community-based and social services programme evaluation and evaluation research, cultural contexts, and multicultural health behaviour. She has worked in community-based settings with cultural communities, using theory-driven evaluation as well as participatory evaluation approaches such as democratic, transformative and empowerment evaluation. Dr. Bledsoe is the author of chapters and articles

focusing on cultural competence and evaluation practice featured in the *American Journal of Evaluation,* the *Journal of Cultural Diversity: an Interdisciplinary Journal,* and work in the volumes, *Evaluation in Action: Interviews with Expert Evaluators*, the *Handbook of Mixed Methods Research, the Handbook of Ethics for Research in the Social Sciences,* work in the edited book, *When Research Studies Go Off the Rails: Solutions and Prevention Strategies,* and work forthcoming in the volume*, Qualitative Inquiry in the Practice of Evaluation*. Dr. Bledsoe has been actively involved in the *American Evaluation Association (AEA)* for over 16 years. She was elected to serve a three year term on the AEA Executive Board, and sat on the task force to develop AEA's public statement on cultural competence in evaluation. She is the past co-chair and current programme chair of AEA's *Programme Theory and Theory-Driven Evaluation Topical Interest Group,* and is a past-Chair of the association's Diversity Committee. She is currently serving as co-chair of AEA's Graduate Education Diversity Internship.

DE LA GARZA, Thania is the Adjunct General Director of Evaluation at the *Consejo Nacional de Evaluación de la Política de Desarrollo Social* (CONEVAL) in México, since 2009. She obtained a Master in Public Policy (MPP) degree at the *Instituto Tecnológico Autónomo de México* (ITAM) and has a Bachelor degree in Economics from the same university. From 2000 to 2007 she worked at the National Institute of Social Development (INDESOL) as Director of Research and Professionalization and, subsequently, held the post of Director of the Evaluation Unit at the same institution. Furthermore, she has taught programme evaluation, public administration and evaluation of projects at different universities in México, (*Instituto Tecnológico Autónomo de México* (ITAM), *Universidad Iberoamericana* (UIA) and *Universidad Nacional Autónoma de México* (UNAM).

DE SILVA, Soma has worked for nearly twenty years in the field of evaluation commissioning, organizing and supporting evaluations of national development programmes. She is a former Regional Adviser, Monitoring and Evaluation for the UNICEF Regional Office for South Asia. She has served as the Head of the Population Census Division of the Department of Census and Statistics in Sri Lanka where she managed national demographic surveys and research and population census activities. She is a founder member and two times past President of the Sri Lanka

Evaluation Association and a founder member of the United Nations Evaluation group for the Asia Pacific (UNEDAP). She is currently serving as the President of the International Organization for Cooperation in Evaluation (IOCE). She initiated and coordinated the South Asia Evaluation volume published in 2008. She has initiated and is currently coordinating a project sponsored by the International Development Research Centre: 'Teaching Evaluation in South Asia' which is designed to establish evaluation training at post graduate diploma level in academic institutions in South Asia.

ERIKSEN, Janie is a Senior Evaluation Specialist in the Evaluation Office in UNICEF New York managing programme performance assessment and institutional effectiveness portfolios. She previously worked with UNHCR, Danida and as Alternate Executive Director in the African Development Bank and in consultancy work *inter alia* on the UN Evaluation of the Paris Declaration on Aid Effectiveness and on the UN Delivering as One evaluability study. She has been a member of the UNEG Human Rights and Gender Equality task force since 2009.

GARCIA, Oscar A. is a Senior Evaluation Adviser at the Evaluation Office of the United Nations Development Programme (UNDP). He oversaw programmatic evaluations for Africa and Latin America and the Caribbean. He co-chaired the UNEG task forces on Country Level Evaluations, Capacity Development in Evaluation, and Harmonization of Evaluations in the UN system. Prior to joining UNDP he was founder and managing director of Proactiva, a consulting firm specializing in strategic management and development evaluation. He held positions in public administration in Bolivia, where he was Director General for Trade Policies and resource mobilization manager for the Social Emergency Fund. He consulted throughout Latin America for IADB, WB, UNDP, and UNFPA. He is author of several evaluation reports and papers on development evaluation.

GREENE, Jennifer C. is a professor of Educational Psychology at the University of Illinois at Urbana-Champaign. Her work focuses on the intersection of social science methodology and social policy and aspires to be both methodologically innovative and socially responsible. Greene's methodological research has concentrated on advancing qualitative and mixed methods approaches to social inquiry, as well as democratic com-

mitments in evaluation practice. Greene has held leadership positions in the American Evaluation Association and the American Educational Research Association. She has also provided editorial service to both communities, including a six-year position as co-editor-in-chief of *New Directions for Evaluation*. Her own publication record includes a co-editorship of the recent *Sage Handbook of Programme Evaluation* and authorship of *Mixed Methods in Social Inquiry*. Greene is the 2011 president of the American Evaluation Association.

GUZMAN, Francisco L. has worked in international development for 32 years, in Latin America, Africa, Asia and the Arab States. He has been in the field of evaluation since 1988. He is currently a Senior Evaluation Officer at the International Labour Organization (ILO) which he joined in 2007. Prior to coming to the ILO he worked for the Inter-American Development Bank in the Office of Evaluation and Oversight (OVE) since 1988 and before that for United States Peace Corps in various capacities including Country Director and Deputy Country Director in Haiti and Costa Rica. He has conducted oversight reviews of corporate strategies, processes and instruments; provided normative guidance on evaluation issues; and contributed to evaluation capacity building in Latin America and the Caribbean. He has led numerous evaluation teams and authored numerous ex-post project evaluations, and high-level strategy and country programme evaluations.

HAY, Katherine is a Senior Programme Specialist, Evaluation, in the International Development Research Centre (IDRC). Katherine is based in IDRC's Regional Office in New Delhi, India. Working in international development evaluation, Katherine has domain expertise and interest in social exclusion, gender equality, and capacity strengthening. A champion of evaluation field building, her work includes building evaluation curriculum in universities in South Asia and supporting evaluation communities of practice. Katherine's work and ideas on *Evaluation Field Building in South Asia* were the focus of a Forum in the American Journal of Evaluation (2010). Katherine was instrumental in the conceptualization and realization of the first ever *Evaluation Conclave* held in New Delhi in October 2010 (www.evaluationconclave.org). She has seeded and supported a growing group of South Asian evaluators through a group called the *Community of Evaluators* (www.communityofevalutors.org). Katherine's work on policy

research and social sciences in South Asia is captured in "Making Research Matter" in *Economics and Political Weekly* (Jan 2010). She co-edited a book and CD project with Navsharan Singh called *Real Rights: Decentralization and Women in South Asia* in 2010. She produced a video, book, and CD project on *Women's Empowerment: Research and Programming* in 2009. She is currently editing, with Shubh Kumar Range, a book on *Evaluation in South Asia* (title TBC) and has edited with Ratna Sudarshan a special issue of the Indian Journal of Gender Studies on *Development, Gender, and Evaluation – Reflecting on Practice, Strengthening Practice.* Katherine has been researching, working, and living in South Asia for over 15 years.

HERNÁNDEZ, Gonzalo is the Executive Secretary of the National Council for the Evaluation of Social Policies (Coneval) since November 2005. He directs and coordinates the activities of Coneval, which aims to evaluate politics and social development programmes as well as to measure poverty in the country. Hernández has a Ph.D. in Economics from Oxford University, in England; a MA in Economics from Essex University, England; and a BSc in Economics from ITAM. (*Instituto Tecnológico Autónomo de México*). He was General Director of Evaluation and Monitoring at Sedesol (*Secretaría de Desarrollo Social*) from August 2002 to November 2005. Currently and since 1998 he is member of the Editorial Board of the journal *Economía Mexicana*, published by the Center for Economic Research and Teaching. Also, since 1997, he is member of the National System of Researchers. He was full-time professor at the Economics Department of ITAM from 1991 to 1992 and from 1996 to 2002. He was Director of the Undergraduate Programme of Economics in the same institution from 1998 and 1999. From 1996 to 2000 he was the Academic Representative for the Commission for Labor Cooperation in the Agreement on Labor Cooperation in North America.

HOPSON, Rodney K. is Professor, Department of Educational Foundations and Leadership in the School of Education, and faculty member in the Center for Interpretive and Qualitative Research, in the School of Liberal Arts, Duquesne University. He received his Ph.D. from the Curry School of Education, University of Virginia and has done post-doctoral/sabbatical studies in the Faculty of Education, University of Namibia, the Johns Hopkins Bloomberg School of Public Health and the Centre

of African Studies, Cambridge University. Hopson's research interests lie in social politics and policies, foundations of education, sociolinguistics, ethnography, and evaluation. Relative to his varied research interests, Hopson raises questions that 1) analyse and address the differential impact of education and schooling on marginalized and under-represented groups in the United States, southern Africa, and other nation states that attempt to come to terms with democracy, social change, and equity and 2) seek solutions to social and educational conditions in the form of alternative paradigms, epistemologies, and methods for the way the oppressed and marginalized succeed in global societies.

HUMMELBRUNNER, Richard is Senior Associate of ÖAR Regionalberatung Graz, Austria with more than 30 years of professional experience as a consultant/ evaluator in the fields of regional and international development. During recent years he has been active in promoting the use of systems thinking in evaluation as a practitioner, trainer and author.

KIRKHART, Karen E. holds a Ph.D. in Social Work and Psychology from the University of Michigan and is currently Professor, School of Social Work, David B. Falk College of Sport and Human Dynamics, Syracuse University. She has been active in the evaluation profession for over thirty years. In 2007, her contributions were recognized by the American Evaluation Association (AEA) with the Paul F. Lazarsfeld Award for Outstanding Contribution to Evaluation Theory and the Robert Ingle Award for Outstanding Services to the AEA. Her work addresses culture, validity, and evaluation influence. She was a member of the writing team for the AEA Public Statement on Cultural Competence in Evaluation and collaborated with Carrie Billy, Joan LaFrance and Richard Nichols in their development of an Indigenous Evaluation Framework.

KUSHNER, Saville is a Professor of Public Evaluation at the University of the West of England, UK. He is a prominent theorist and practitioner of programme and policy evaluation, having conducted and directed evaluation projects for more than 30 years in Britain and overseas. He has published numerous books and articles on evaluation, including in the field of international development. Between 2005 and 2007 Saville served as Regional Monitoring & Evaluation Officer for UNICEF in Latin Amer-

ica and the Caribbean, and since then has worked continuously as a UNICEF consultant in Asia and Latin America. His contribution to evaluation theory came with his book entitled *Personalising Evaluation* (Sage, 2000).

MERTENS, Donna M. is a Professor in the Department of Educational Foundations and Research at Gallaudet University, where she teaches advanced research methods and programme evaluation to deaf and hearing students. She also serves as editor for the Journal of Mixed Methods Research. The primary focus of her work is transformative mixed-methods inquiry in diverse communities that prioritizes ethical implications of research in pursuit of social justice. Her recent books include *Programme Evaluation Theory to Practice: A Comprehensive Guide*, *Transformative Research and Evaluation*, *The Handbook of Social Research Ethics*; *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods (3rd ed.)*; *Research and Evaluation Methods in Special Education*; and *Parents and Their Deaf Children: The Early Years*. She is widely published in the *Journal of Mixed Methods Research*, *American Journal of Evaluation*, *American Annals of the Deaf*, and *Educational Evaluation and Policy Analysis*.

MUIR, Janice is a project leader in the Inspection and Evaluation Division of the United Nations Office of Internal Oversight Services. (OIOS) Prior to joining the UN she worked in public policy research and analysis in a number of Australian government departments and agencies. She has been a member of the UNEG Human Rights and Gender Equality task force since 2009.

PAREDES, María Fernanda is currently part of the team of the General Direction of Evaluation at the National Council for the Evaluation of Social Policies (CONEVAL) where she has worked since 2010. She provides technical assistance and develops policy analysis to support the General Director´s Office. She also collaborates as Assistant Professor in the Social Development Specialization at *Universidad Nacional Autónoma de México* (UNAM) and in the Logical Framework Methodology Course at *Instituto Tecnológico Autónomo de México* (ITAM). She worked as a research assistant in a Mexican Think Tank, *Centro de Investi-*

gación para el Desarrollo, A.C. (CIDAC), in which she carried out analysis for governmental institutions in the areas of strategic planning, capacity building, performance evaluation, audit, budget transparency and accounting approval at state and municipal levels. She studied International Affairs at ITAM.

PATTON, Michael Quinn is an independent organizational development and evaluation consultant. He is former President of the American Evaluation Association and author of several evaluation books including *Essentials of Utilization-Focused Evaluation* (Sage, 2012) and *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (Guilford Press, 2011). He teaches regularly in *The Evaluators' Institute*, The World Bank's *International Programme in Development Evaluation Training* in Ottawa, and the American Evaluation Association's professional development courses. He was on the social sciences faculty of the University of Minnesota for 18 years, including five years as Director of the Minnesota Center for Social Research.

REDDY, Shravanti is an Evaluation Specialist in the UN Women Evaluation Office, managing corporate evaluations of gender equality programming and policies and providing technical support to decentralized evaluation processes. She is currently the co-Chair of the UNEG Taskforce on Human Rights and Gender Equality. Prior to joining UNIFEM/UN Women, Ms. Reddy worked on UNICEF's Gender Mainstreaming Assessment and with various civil society organizations in India, the United States and Zimbabwe. She holds a Masters in International Affairs (MIA) from Columbia University's School of International and Public Affairs.

REYNOLDS, Martin is a consultant, researcher, and lecturer in Systems Thinking at The Open University, UK. He has produced distance learning resources for postgraduate programmes on International Development, Environmental Decision Making, and Systems Thinking in Practice and has published widely in these fields. Many of his publications are available on open access. He provides workshop support and facilitation for professional development in systems practice and critical systems thinking and has experience with designing and facilitating workshops for the European Evaluation Society and the American Evaluation Association.

ROGERS, Patricia is Professor of Public Sector Evaluation at RMIT University (Royal Melbourne Institute of Technology) in Australia. She has worked in public sector evaluation and research for more than 25 years, with government and non-government organizations (international, national, state and local) across a wide range of programme areas, including agriculture, community development, criminal justice, early childhood, education, health promotion, Indigenous housing, and legal aid. She has worked on projects with international agencies including the United Nations Development Programme, the World Bank Institute, the Network of Networks on Impact Evaluation (NONIE), and Inter-Action. She is leading the development of BetterEvaluation, a collaboration to improve evaluation by sharing information about methods. Her research interests focus on appropriate methodological choices to suit the purpose of evaluation and the nature of what is being evaluated, including attention to complication and complexity. Her recent publications include *Purposeful Programme Theory: Effective Use of Theories of Change and Logic Models*, (Jossey-Bass/Wiley, 2011) with Sue Funnell.

RUGH, Jim has been professionally involved for 47 years in rural community development in Africa, Asia, Appalachia and other parts of the world. For the past 31 years he has specialized in international programme evaluation. He served as head of Design, Monitoring and Evaluation for Accountability and Learning for CARE International for 12 years. He has also evaluated and provided advice for strengthening the M&E systems of a number of other International NGOs. He is recognized as a leader in the international evaluation profession. He serves as the American Evaluation Association (AEA) Representative to and Vice President of the International Organization for Cooperation in Evaluation (IOCE), the global umbrella of national and regional professional evaluation associations. Jim co-authored the popular and practical *RealWorld Evaluation* book (published by Sage in 2006) and has led numerous workshops on that topic for many organizations and networks in many countries. In recognition of his contribution to the evaluation profession he was awarded the 2010 Alva and Gunnar Myrdal Practice Award by AEA.

SEGONE, Marco is responsible for the Decentralized evaluation function as well as the National evaluation capacity development portfolios at UNICEF Evaluation Office. He has co-chaired the UNEG Task Force on National evaluation capacities since 2009 and serves as Senior Advisor to the IOCE Board. Previously, he was Regional Chief, Monitoring and Evaluation in the UNICEF Regional Office for Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS), during which time he represented UNICEF on the Board of Trustees of the International Programme Evaluation Network (IPEN). During his 21 years in international development, he has worked in Bangladesh, Pakistan, Thailand, Uganda and Albania in integrated development projects. In 1996 he joined UNICEF to work for the UNICEF Regional Office for Latin America and the Caribbean; UNICEF Niger, where he founded and for two years coordinated the Niger M&E Network (ReNSE); and UNICEF Brazil, where he was one of the founders and coordinator of the Brazilian Evaluation Network. In 2003 he was elected Vice-President of IOCE and was one of the founders of the Latin America and the Caribbean Network for Monitoring, Evaluation and Systematization (RELAC). Segone authored and/or edited several books and articles, including: *From policies to results*; *Country-led M&E systems*; *Bridging the gap. The role of M&E in evidence-based policy making*; *New trends in development evaluation*; *Creating and developing evaluation organizations*; and *Democratic evaluation.*

UITTO, Juha I. is Deputy Director of the United Nations Development Programme (UNDP) Evaluation Office. He joined the office as Evaluation Adviser in 2006 and has conducted and managed several programmatic and thematic evaluations, particularly in the Asia and the Pacific region. He also serves as the Executive Coordinator of the UN Evaluation Group (UNEG). In the past, Juha has worked as evaluator, researcher and consultant in international development with the Global Environment Facility (GEF), the International Fund for Agricultural Development (IFAD) and other agencies. He was for nine years on the faculty of the United Nations University. A geographer by training, he holds an MSc from the University of Helsinki in his native Finland and a PhD in Social and Economic Geography from Lund University in Sweden. He has authored a large number of articles, book chapters and reports on topics related to environment and development, natural resources management, environmental hazards, and evaluation.

VALDEZ, Brenda holds a B.A. degree in Economics and a Specialization degree in Applied Econometrics from the *Universidad National* of Mexico (UNAM). She has been part of the impact evaluation team of the National Council for the Evaluation of Social Development Policy (CONEVAL) in México since 2010. Among her work responsibilities she is in charge of reviewing and analyzing the impact evaluations of the federal social programmes. Additionally, the impact evaluation team also provides technical guidance to the federal dependencies that are planning to implement impact evaluations for their social programmes. Since 2010 she has combined her work at CONEVAL with teaching assistance activities in Political Economy I and II lectures at UNAM.

WILLIAMS, Bob is an evaluation consultant based in New Zealand. He works all over the world, largely promoting the use of systems ideas in the evaluation field. Bob is co-author, with Richard Hummelbrunner, of *Systems Concepts in Action: A Practitioner's Toolkit* (2010); an associate author of *Making Evaluations Matter: A Practical Guide for Evaluators* (2011), and a member of three evaluation journal editorial boards. He is an experienced facilitator and workshop designer. In the New Zealand winter months Bob runs a programme for skiers and snowboarders with physical and intellectual disabilities.

Abbreviations

ACWF	All China Women's Federation
ADR	Assessment of Development Results
BA	Beneficiary Assessment
BIA	Benefit Incidence Analysis
CCAS	Conference Committee on the Application of Standards
CDI	National Commission for the Development of the Indigenous Population (Mexico)
CEACR	Committee of Experts on the Application of Conventions and Recommendations
CEDAW	Convention on the Elimination of All Forms of Discrimination Against Women
CIDA	Canadian International Development Agency
CHAT	Cultural Historical Activity Theory
CONEVAL	Mexico's National Council for the Evaluation of Social Policies
CRE	Culturally Responsive Evaluation
CRC	Convention on the Rights of the Child
CSH	Critical Systems Heuristics
CSO	Civil Society Organization
DHS	Demographic and Health Survey
DWCP	Decent Work Country Programme
EAA	European Aid Agency
EED	Specific Performance Evaluation (Mexico)
GE	Gender Equality
HDI	Human Development Index

HR	Human Rights
IDEAS	International Development Evaluation Association
IDRC	International Development Research Centre
ILC	International Labour Conference
ILO	International Labour Organization
ILS	International Labour Standards
IMSS	Mexican Social Security Institute
INEGI	National Institute of Statistics and Geography (Mexico)
INTRAC	International Evaluation Conference
IOCE	International Organization for Cooperation in Evaluation
IPDET	International Programme for Development Evaluation Training
IPEN	International Programme Evaluation Network
ITC	International Training Centre
JCSEE	Joint Committee on Standards for Educational Evaluation
KAP	Knowledge, Attitude and Practice
LSMS	Living Standards Measurement Survey
M&E	Monitoring and Evaluation
MDGs	Millennium Development Goals
MGNREGS	Mahatma Gandhi National Rural Employment Guarantee Scheme
MICS	Multiple Indicators Cluster Survey
MOCA	Ministry of Civil Affairs
MoE	Ministry of Education

MoV	Means of Verification
NDRC	National Development and Reform Council
NED	Non-experimental designs
OBW	Outcome-based Workplans
OECD-DAC	Development Assistance Committee of the Organization for Cooperation and Development
OVI	Objectively Verifiable Indicator
P&B	Programme and Budget
PETS	Public Expenditure Tracking Studies
PGA	Participatory Gender Audit
PPA	Participatory Poverty Assessment
PPCMJ	Promocion de Convenios en Materia Justica (Mexican programme for the rights of the indigenous population)
QUAL	Qualitative methods
QUANT	Quantitative methods
RBM	Results-based management
RBSA	Regular Budget Supplementary Account
RCT	Randomized Controlled Trial
SEDESOL	Ministry of Social Development (Mexico)
SEP	Ministry of Public Education (Mexico)
SPF	Strategic Policy Framework
SSM	Soft Systems Methodology
TC	Technical Cooperation
UN ECOSOC	United Nations Economic and Social Council
UNEG	United Nations Evaluation Group

UNESCO	United Nations Education, Scientific and Cultural Organization
UNDAF	United Nations Development Assistance Framework
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
UNIFEM	United Nations Development Fund for Women (now replaced by UN Women)
UN-NGLS	United Nations Non-Governmental Liaison Service
UN Women	United Nations Entity for Gender Equality and the Empowerment of Women
WFP	World Food Programme
YEM	Youth Employment and Migration

Learn how to design and manage Equity-focused evaluations!

Download the manual, free of charge, at :
www.mymande.org/selected-books

Interact live with world-level evaluators!

Take part at the live Webinars available, free of charge, at www.mymande.org

To participate, you need a computer and an internet connection.

Access hundreds of resource material and be part of a global evaluation community!

At MyM&E, an interactive web 2.0 platform, you can, free of charge:

- download hundreds of evaluation manuals and material;

- be part of a global evaluation community by developing your social profile and networking;

- participate and share your knowledge through blogs;

- watch videos of keynote speakers;

- find technical advice on how to design and manage an evaluation in the practical "How to" section;

- Search in the inventory of training delivered by different institutions all over the world;

- Post your CV or look for a consultant in international evaluation rosters;

- and much more!

All of this, available at www.mymande.org/webinars

Read good practices and lessons learned about Country-led monitoring and evaluation systems!

Download, free of charge, the book with articles by almost 30 authors from different institutions, at: www.mymande.org/selected-books, and then click on virtual library

unicef

Country-led monitoring and evaluation systems
Better evidence, better policies, better development results

It contains DevInfo databases

In partnership with:

DevInfo IDEAS IOCE

MICS UNECE THE WORLD BANK
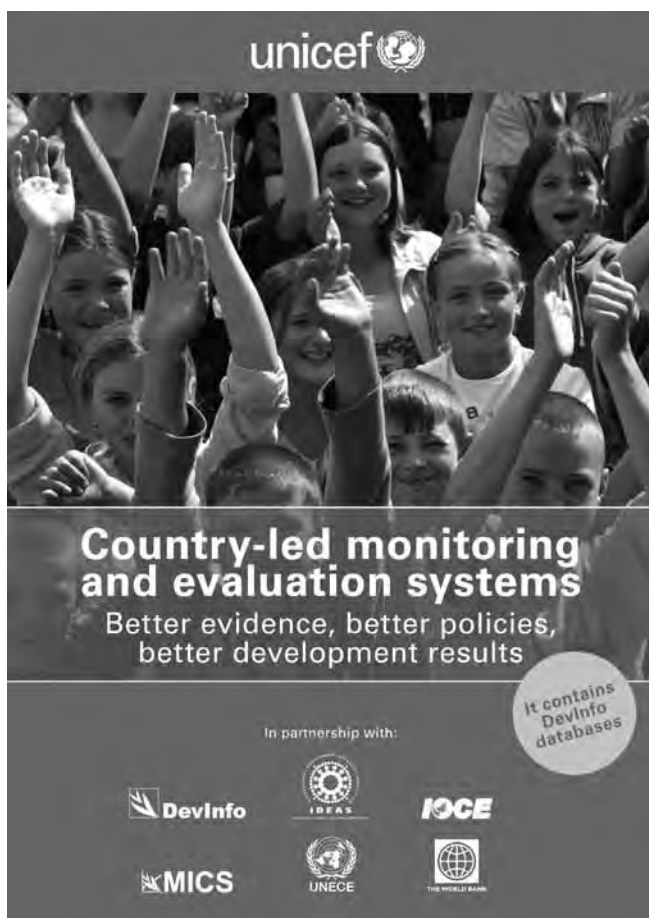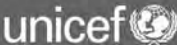
Read how Evaluation can and should contribute to policy making!

Download, free of charge, the book with articles by almost 30 authors from different institutions, at: www.mymande.org/selected-books, and then click on virtual library

unicef

Bridging the gap
The role of monitoring and evaluation in evidence-based policy making

In partnership with:

DevInfo   IDEAS   MICS   THE WORLD BANK
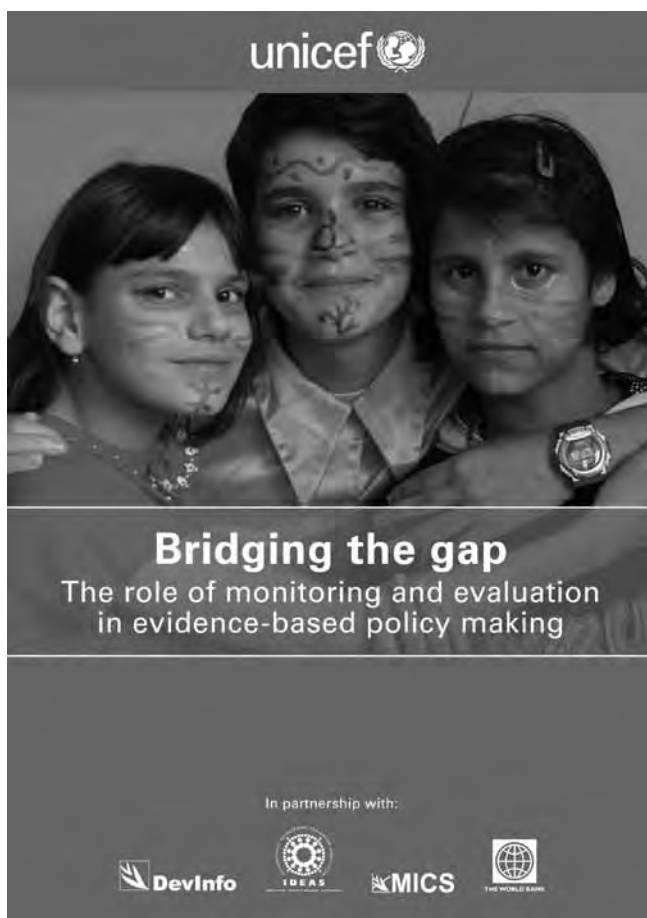
Learn how to strengthen national evaluation capacities!

Download, free of charge, the book with articles by almost 30 authors from different institutions, at: www.mymande.org/selected-books, and then click on virtual library

unicef

From policies to results
Developing capacities for country monitoring and evaluation systems

In partnership with:

DevInfo IDEAS International Labour Office IOCE

THE WORLD BANK UNDP UNIFEM WFP wfp.org

It is high time to put equity at the center of efforts to promote development. Addressing a high level meeting on the Millennium Development Goals in Tokyo last year, UNICEF's Executive Director, Tony Lake, put the point eloquently. He declared: "There can be no true progress in human development unless its benefits are shared – and to some degree driven – by the most vulnerable among us... the equity approach is not only right in principle. It is right in practice".

In the same vein, it is an appropriate moment to ask whether evaluation as a discipline and evaluators as a profession are addressing equity issues in ways which are indeed right in principle and right in practice. Some of the answers can be found in the present volume, which brings together a tremendous richness and diversity of evaluation thinking and experience. While a number of the papers included in the collection touch on approaches and methods already familiar to evaluators, the challenge of addressing the question of equity has helped to demonstrate renewed relevance and establish fresh perspectives. Several essays showcase examples of evaluations addressing equity issues, providing a valuable source of inspiration.