



How to design and manage
Equity-focused evaluations



How to design and manage
Equity-focused evaluations

The Evaluation Working Papers (EWP) are documents that present strategic evaluation findings, lessons learned and innovative approaches and methodologies. We would like to encourage proposals for relevant papers which could be published in the next EWP issues. Papers can be prepared by UN staff and by partners.

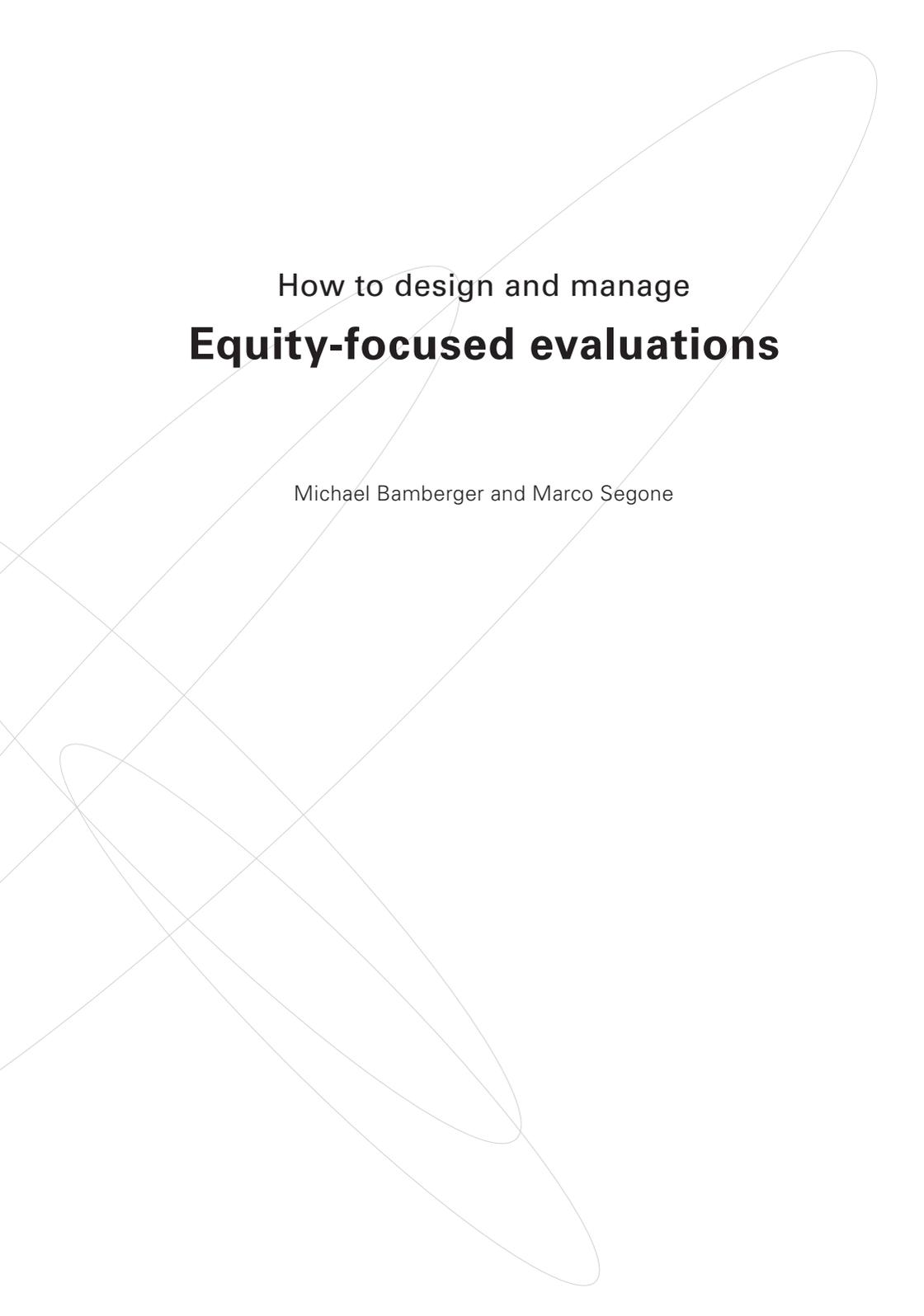
For additional information and details please contact Marco Segone, Systemic strengthening, UNICEF Evaluation Office, msegone@unicef.org

- ISSUE # 1: New trends in development evaluation. Published jointly with 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.

Disclaimer:

The opinions expressed are the personal thinking of the contributors and do not necessarily reflect the policies or views of UNICEF or any other organization involved or named in this publication. The text has not been edited to official publication standards and UNICEF and partner organizations accept no responsibility for errors.

Extracts from this publication may be freely reproduced with due acknowledgement.



How to design and manage
Equity-focused evaluations

Michael Bamberger and Marco Segone

Contents

Editorial..... **X**

Acknowledgements..... **XII**

Part 1

Equity and Equity-focused evaluations

Section 1. What is equity and why does it matter? 2

- 1.1 The challenge of achieving equitable development results for children 2
- 1.2 What is equity? 3
- 1.3 Why does equity matter? 4
- 1.4 Why is equity now so urgent ? 7
- 1.5 What are the implications for the evaluation function? 8

Section 2. Defining Equity-focused evaluations 9

- 2.1 What is an Equity-focused evaluation? 9
- 2.2 Why are Equity-focused evaluations needed? 9
- 2.3 Purposes of Equity-focused evaluations 10
- 2.4 Empowering worst-off groups, including children, through Equity-focused evaluation processes 12

Part 2

Managing Equity-focused evaluations

Section 3. Preparing for the evaluation 18

- 3.1 Determining the evaluability of the equity dimensions of the intervention 19
- 3.2 Identifying evaluation stakeholders, including worst-off groups ... 26
- 3.3 Identifying the intended use and intended users 27
- 3.4 Identifying potential challenges in promoting and implementing Equity-focused evaluations 29

Section 4. Preparing the evaluation Terms of Reference 33

- 4.1 Defining the scope and purpose of the evaluation 33
- 4.2 Framing the evaluation questions 33
- 4.3 Selecting a technically-strong and culturally-sensitive evaluation team 39

Section 5. Designing the evaluation	41
5.1 Selecting the appropriate evaluation framework	41
A. Theory-based Equity-focused evaluation	41
<i>Basic components of a programme theory of change</i>	43
<i>Refinements to the basic logic model</i>	45
B. The bottleneck analysis framework.....	45
<i>Use of services by worst-off groups</i>	46
<i>Supply side factors</i>	49
<i>Demand side factors</i>	49
<i>Contextual factors</i>	50
5.2 Selecting the appropriate evaluation design.....	50
A. Mixed methods designs	51
<i>Types of mixed method designs</i>	52
<i>Triangulation: a powerful tool for assessing validity and for deepening understanding</i>	56
B. Attribution, contribution and the importance of the counterfactual	58
C. Equity-focused evaluation at the policy level.....	61
<i>Systems approaches to evaluation</i>	62
<i>Unpacking complex policies</i>	66
<i>Pipeline designs</i>	66
<i>Policy gap analysis</i>	67
<i>Using other countries or sectors as the comparison group</i>	68
<i>Concept mapping</i>	69
<i>Portfolio analysis</i>	69
D. Equity-focused evaluations at the project and programme levels	70
<i>Conventional quantitative impact evaluation designs</i>	70
<i>Estimating project impacts using non-experimental designs (NEDs)</i>	72
<i>Potentially strong non-experimental designs</i>	73
E. Feasibility analysis	74

5.3	Collecting and analyzing data	74
A.	Collecting data and analyzing contextual factors	75
B.	Collecting and analyzing surveys and qualitative information to understand knowledge, attitude and practice	78
C.	Collecting and analyzing information on the quality of services delivered and the satisfaction of citizens	81
	<i>Citizen report cards</i>	81
D.	Carrying out cost-effectiveness studies to compare costs and results of alternative intervention	82
	<i>Cost-effectiveness analysis</i>	82
	<i>Public expenditure tracking (PETS) studies</i>	83
	<i>Public expenditure Benefit Incidence Analysis (BIA)</i>	84
Section 6.	Utilizing the evaluation	86
6.1	Preparing the evaluation report and alternative forms of reporting	86
6.2	Disseminating the evaluation and preparing a Management Response	87
Section 7.	Conducting Equity-focused evaluations under real-world constraints	89
7.1	Understanding the evaluation scenario	89
7.2	Reconstructing the programme theory when it is non-existent or is very weak.....	90
7.3	Conducting credible Equity-focused evaluations when working under budget and time constraints	92
7.4	Reconstructing baseline data when the evaluation is not commissioned until late in the implementation cycle.....	94
7.5	Conducting Equity-focused evaluations in countries where government is not supportive of an equity focus	96
Section 8.	Case studies of UNICEF-supported Equity-focused evaluations	98
References	102



Editorial & Acknowledgements

Editorial.....	X
Acknowledgements.....	XII

EDITORIAL

The push for a stronger focus on equity in human development is gathering momentum at the international level. Its premise is increasingly supported by United Nations reports and strategies as well as by independent analysis and donors. More and more national policies and international alliances are focusing on achieving equitable development results for children. While this is the right way to go, it poses important challenges – and opportunities – for the evaluation function. How can one strengthen the capacity of Governments, organizations and communities to evaluate the effect of interventions on equitable outcomes for children? What are the evaluation questions to ensure interventions are relevant and are having an impact in decreasing inequity, are achieving equitable results, and are efficient and sustainable? What are the methodological implications in designing, conducting, managing and using Equity-focused evaluations? This document represents a first attempt to address these questions.

The purpose of this document is to provide guidance to UNICEF Country Offices, their partners and Governmental and Civil Society stakeholders, including communities and worst-off groups, on how to design and manage evaluations to assess the contribution of policies, programmes and projects to equitable development results for children. The document is complemented by a resource centre on Equity-focused evaluation available at www.mymande.org, in which readers can access in more detail the methodological material presented in this document. Therefore, this document should be used together with that electronic resource centre.

This document is divided into two parts. Part I discusses **Equity and Equity-focused evaluations**. Section 1 defines equity, why equity matters and why equity is so urgent now. Section 2 defines Equity-focused evaluations, explaining what their purpose should be and potential challenges in their promotion and implementation.

Part II explains **how to manage Equity-focused evaluations**, explaining the key issues to take into account when preparing for Equity-focused evaluations (section 3) and developing the Terms of Reference (section 4); designing the evaluation, including identifying the appropriate evaluation framework, evaluation design and appropriate methods to collect data (section 5); and utilizing the evaluation (section 6). Section 7 explains how to conduct Equity-focused evaluations under real-world constraints.

Eight case studies are included in section 8 to illustrate how evaluations supported by UNICEF have addressed equity-focused issues in a post-conflict education project in Timor l'Este; an education project in Nepal; a community sanitation project in Cambodia; the humanitarian response to a population displacement crisis in Pakistan; a community schools project in Egypt; a community justice facilitation project in Tanzania; the child protection component of the tsunami emergency programme in Indonesia; and a programme to reduce female circumcision in Senegal. A ninth case study shows how policy-gap analysis was applied to assess the impacts of social assistance on reducing child poverty and child social exclusion in Albania.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Colin Kirk, Director, UNICEF's Evaluation Office for his constant encouragement and support in making this document possible.

This document benefitted from the inputs of Senior Managers working at UNICEF both at Headquarters, Regional Offices and Country Offices. At Headquarter, we would like to acknowledge Evaluation Office specialists (Samuel Bickel; Krishna Belbase; Robert McCouch; Janie Eriksen; Kathleen Letshabo; Erica Mattellone; Abigail Taylor; Ashley Wax; and Chelsey Wickmark), as well as senior colleagues in several departments (Maniza Zaman, Deputy Director, Survival, Growth and Development Programmes; Robert Jenkins, Associate Director, Policy and Practice; Mickey Chopra, Associate Director, Child Survival; Susan Durston, Associate Director, Education; Karin Heissler, Head of Capacity Building Cluster in Child Protection; Jennifer Kean, Child Protection Specialist; Abdelmajid Tibouti, Senior Adviser, Programme Division; Attila Hancioglu, Global MICS Coordinator; and David Anthony, Chief of Policy Advocacy). At regional level, valuable feedback and comments were provided by the Regional Monitoring and Evaluation Chiefs: Ada Ocampo from the Asia Regions (APSSC); Anne-Claire Luzot from Central and Eastern Europe and the Commonwealth of Independent States Region; Bastiaan Van't Hoff, Enrique Delamonica and Lucio Goncalves from the Americas and the Caribbean Region; Edward Addai, Dorothy Rodzga, Pedersen Bo and Vincenzo Vinci from East and South Africa Region; Christina Bierring, Herve Peries and Pascale Kervyn from West and Central Africa Region; and Pierre Ngom from Middle East and North Africa Region. At Country Office level: Hicham Mansour (Morocco); Marjan Montazemi (Iran); Yahaya Balima, Toshi Yamamoto and Mohammed Youssouf Abdallah (Djibouti); Jane Mwangi and Julianna Lindsay (Ghana); Maria Antonia Scolarmiero and Marcio Carvalho (Brazil); and Ennio Cufino (Argentina).

The content of this document also benefitted from discussions in sessions on Equity-focused evaluations at the UNICEF Global Monitoring and Evaluation (M&E) Meeting in New York, held in April 2011, as well as Regional M&E Workshops for the Middle East and North Africa Region in Amman, Jordan, in July 2011 and for South Asia in Colombo, Sri Lanka, in June 2011. Contributions discussed at the Global IDEAS Conference in Amman, Jordan, and at the Conference of the Sri Lanka Evaluation Association, also added important perspectives external to UNICEF.

As Equity-focused evaluation is an emerging area of work, the authors welcome feedback from readers. Please direct your comments to: Marco Segone at msegone@unicef.org and Michael Bamberger at jmichaelbamberger@gmail.com



Part 1

Equity and

Equity-focused

evaluations

Section 1. What is equity and why does it matter?	2
Section 2. Defining Equity-focused evaluations	9

SECTION 1: WHAT IS EQUITY AND WHY DOES IT MATTER?

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

1.1 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. In his foreword to the *Millennium Development Goals Report 2010*, United Nations Secretary-General Ban Ki-Moon argues that "the world possesses the resources and knowledge to ensure that even the poorest countries, and others held back by disease, geographic isolation or civil strife, can be empowered to achieve the MDGs." That report underscores the commitment by the United Nations and others to apply those resources and that knowledge to the countries, communities, children and families who are most in need (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 inequities. 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).

1.2 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 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 its cause is due to the social context, rather than to biological fac-

tors. For example, young adults tend to be healthier than elderly adults, and female newborns generally have lower birth weights than male newborns. These disparities cannot be described as inequities since they are caused by unavoidable biological factors. If, however, girls and boys showed dramatic differences in nutritional status or immunization levels, the disparity would likely be due to social rather than biological factors, and would therefore be considered unnecessary and avoidable. Gender discrimination and other social, political, and economic forces that systematically deny the rights of specific groups – such as girls, children of minority groups, or children with disabilities – are cause for grave concern from an equity perspective.

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 that include but are by no means limited to: gender discrimination; ethnic, linguistic, minority, and religious discrimination; discrimination due to disability status; structural poverty; natural or man-made disasters; geographic isolation; cultural and social norms; and weak governance.

An equity-focused intervention must therefore begin with an analysis of the context in which inequity operates. This analysis informs the design of programmes 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.

1.3 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 the 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. Immunizations with pneumococcal conjugate and rotavirus vaccines have the potential to accelerate progress towards reducing pneumonia and diarrhea, among the foremost killers of poor children. Recently developed interventions such as mother-baby packs of anti-retroviral medicines have the potential to expand access to the many women and children still missing out on vital services to combat HIV and AIDS. The spread of SMS (Short Message Service) technology is allowing more data to be collected rapidly, enabling improved targeting of interventions to those most in need.

...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 economic growth of a country, and so higher rates of economic 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. Equity is a cornerstone of effective and harmonious relationships between people and it underpins our common systems of ethics and rights (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 inequality 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 inequality 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 school. This vicious cycle reinforces the inequity in education impacting negatively on income inequality, 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 their persistence, or worsening, over the decades makes inequity become accepted as “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.

1.4 Why is equity now so urgent?

In addition to the reasons explained above, equity is becoming urgent due to at least five major global threats that could undermine accelerated progress towards equitable development for children: the food and financial crises; rapid urbanization; climate change and ecosystem degradation; escalating humanitarian crises; and, heightened fiscal austerity (UNICEF, 2011a).

Such global trends, however dire, can also present opportunities for change and renewal – if governments and other stakeholders seize upon these challenges to demonstrate their commitment to equitable development results, including MDGs, and work together to hasten progress towards them. The central focus for meeting the MDGs with equity is clear: the need to prioritize the poorest and worst-off children and families, and to deepen investment for development. The good news is that the push for a stronger focus

on equity in human development is gathering momentum at the international level. Its premise is increasingly supported by United Nations reports and strategies as well as by independent analysis and donors. More and more national policies and international alliances are focusing on achieving equitable development results.

1.5 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 electronic resources mentioned above, represents a first attempt to address these challenges.

SECTION 2: DEFINING EQUITY-FOCUSED EVALUATIONS

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

2.1 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 groups¹. 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 gaps between best-off, average 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.

2.2 Why are Equity-focused evaluations needed?

Equity-focused evaluations look explicitly at the equity dimensions of interventions, going beyond conventional quantitative data to the analysis of behavioral change, complex social processes and attitudes, and collecting information on difficult-to-reach socially marginalized groups. In addition, Equity-focused evaluations pay particular attention to process and contextual analysis, while *con-*

1 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.

*ventional impact evaluation designs*² use a pre-test/post-test comparison group design, which does not study the processes through which interventions are implemented nor the context in which they operate.

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

2.3 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, nature of the intervention 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. In the case of multi-country and/or regional/global evaluations, Equity-focused evaluations can be used to test **key assumptions** of the equity approach in different political, geographical, cultural and social contexts. Some of the assumptions to be tested and questions to be asked could include:

- How effective is the equity approach in different contexts? For example:
 - Comparing effectiveness of the equity approach in poor countries, with a large and dispersed worst-off population,

2 We use the term “conventional impact evaluation design” to refer to designs that use a pre-test/post-test comparison of the group receiving the intervention (e.g. nutritional supplement, improved water supply, community leadership training, anti-corruption programme) with a matched group that does not receive the intervention. The control or comparison group is used as the counterfactual to address the question, “What would have been the condition of the project group if the intervention had not taken place?” Or more simply “what difference did the project make?”

with middle-income countries with a smaller worst-off population.

- Comparing the effectiveness of the equity approach in countries with a strong commitment to equity (and where costs may not be the principle issue), with countries that are less committed to equity and where cost and ease of implementation may be critical.
- Can the equity approach deliver rapid results? What are the “quick-wins”³? What are the strategic areas and sectors where rapid results can and cannot be achieved? Are these results sustainable? What are the main barriers to achieving rapid results?
- What level of results can be achieved within the existing operating procedures of implementing agencies and what results require more fundamental system change?

Evidence-based policy advocacy.⁴ 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

3 The term “quick-wins” is used to refer to impacts that can be achieved easily and economically in a non-controversial way that will demonstrate to partners the benefits of the proposed approaches. This is also referred to as “low-hanging fruit”.

4 Often the term “evidence-based” is used to imply that the only credible evidence is obtained from randomized control trials, with strong quasi-experimental designs (using statistical matching techniques, such as propensity score matching, being a second best approach). However, authors such as Donaldson, Christie and Mark (2009), and Rieper, Leeuw and Ling (2010), underline that different academic disciplines and different stakeholders have different expectations of what is considered “credible” evidence and convincing proof of the effectiveness of an intervention. These and other authors also point out that while many evaluators assume that a statistical counterfactual is considered the only credible way to assess causality, in fact there are many disciplines such as criminal investigation, physics, economic history and public health where it is rarely possible to use a statistically matched comparison group. So the experimental evaluation approach to causality is only one of many approaches and certainly not the norm.

the evidence base, particularly resulting from innovative programming to foster equity, to demonstrate what works in diverse country contexts.

In the specific case of pro-equity interventions, Equity-focused evaluation contributes to two additional main purposes:

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 policy makers 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.

2.4 Empowering worst-off groups, including children, through Equity-focused evaluation processes

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:

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 into 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.

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 relationship 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 under-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 children as appropriate, since children are also among the worst-off groups. The CRC 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.

When considering the participation of children, there is an important distinction between the situation of children who are part of functioning households or social groups (where adults may not wish to involve them in decisions about their welfare), and that of the many children such as child soldiers, victims of trafficking or street children, who may not be part of any functioning household. In the latter case it may be the "experts", government agencies or communities that have decided (explicitly or implicitly) that children cannot speak for themselves. Several different angles can be taken to define the nature of children's participation. Roger Hart (Hart, 1992) used an eight-degree scale, which goes from child-initiated participation to tokenism, embellishment and manipulation of children's opinion by adults. Efforts that fall under tokenism, embellishment and manipulation not only fail in their objective to foster the participation of children, but can also discredit the entire Equity-focused evaluation process and even the organisations involved, ultimately undermining the meaning of the right to participate (UNICEF, 2002).

Context is also important. Political, social and economic contexts will have their own institutional norms and practices at different levels (national, sub-national, community, family), and in different fora will favour (or limit) participation to different degrees. Analysis can reveal how the context limits participation, as well as how participation can be increased. Rakesh Rajani's "Framework for promoting effective adolescent participation" links the above two aspects – context and the relationship between children and adults – with other factors to define the nature of participation. These two frameworks are not only good for designing programmes, but for defining the participatory activities for research and M&E exercises as well, i.e. where children will participate, in what role and through what type of interaction with adults. If the M&E activity itself is designed to build participation, then managers and evaluators must specify how the activity will influence children's capabilities and their supporting environment and therefore their opportunities for participation.



Part 2

Managing Equity-focused evaluations

Section 3. Preparing for the evaluation	18
Section 4. Preparing the evaluation Terms of Reference.....	33
Section 5. Designing the evaluation	41
Section 6. Utilizing the evaluation	86
Section 7. Conducting Equity-focused evaluations under real-world constraints	89
Section 8. Case studies of UNICEF-supported Equity-focused evaluations	98
References.....	102

This document focuses on managing equity-focused evaluations, and gives the technical details necessary to ensure that its implementation is technically rigorous. The term evaluation manager is used throughout to describe the person responsible for organizing and leading the evaluation process, including coordinating its design¹, and who will receive the evaluation report and ensure its quality. The term evaluator/ evaluation team is used to describe the person or team who collects and analyses the data, and prepares the report on findings and recommendations.

SECTION 3: PREPARING FOR THE EVALUATION²

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

Before beginning an evaluation, it is important to assess whether equity dimensions have been adequately considered during the design and implementation of an intervention. This is of fundamental importance because the fulfillment of human rights and avoidance of discrimination are necessary conditions for sustainable development and, therefore, all UNICEF interventions have a mandate to address human rights and equity issues. Thus, UNICEF has an obligation to take these dimensions into consideration when planning an intervention, and beneficiaries of UNICEF interventions also have a right to be engaged in a way that promotes human rights and equity.

It is much easier to evaluate equity dimensions of an intervention when they have been addressed during the design, implementation and monitoring of an intervention. However, the reality is that interventions do not always mainstream human rights and equity. Even if mainstreaming equity is not the focus, it is always important for the evaluation manager and evaluation team to have the skills and knowledge to ensure good assessment of equity dimensions during an evaluation.

- 1 The responsibility of the evaluation manager for the design of the evaluation can range from primary responsibility for developing the design which will be implemented by the evaluation team; to providing general guidelines to the evaluation team who will develop the technical design. When an evaluation is commissioned through an RFP (Request for Proposal) interested evaluators and firms may be required to present very detailed evaluation design proposals.
- 2 This chapter is based on and adapted from *Integrating human rights and gender equality in evaluation*, UNEG, 2011

3.1 Determining the evaluability of the equity dimensions of the intervention³

An *evaluability assessment* is an exercise that helps to identify whether an intervention can be evaluated, and whether an evaluation is justified, feasible and likely to provide useful information. It assesses whether the evaluation can achieve its objectives within the proposed time frame (for example, it may be too early to assess certain kinds of outcomes or impacts), and within the proposed budget and time inputs. Its purpose is not only to decide if the evaluation can be undertaken or not, but also to prepare the intervention to ensure that necessary conditions for an evaluation are in place. The evaluability should also indicate any particular political, social and cultural challenges as well as the technical challenge in conducting an Equity-focused evaluation. This will inform decisions on the level of equity-analysis that the evaluation can realistically cover.

Interventions will generally fall into two categories:

- a. Where equity is the primary focus of the intervention, and
- b. Where equity is not the primary focus of the intervention.

All evaluations in both categories should include an assessment of the equity dimensions of the interventions. For interventions in the first category, equity will be a primary focus of the evaluation. Interventions falling into the second category, where human rights and equity is not the primary focus, will vary in the extent to which equity elements were explicit in the programme design.

Interventions will also differ depending on whether disaggregated information was systematically collected about different groups. In addition, interventions in the second category will differ in their attention to equity during implementation. In both categories, the evaluation methods and procedures for assessing equity dimensions will be similar, although the evaluation questions may differ.

When considering the evaluability of an intervention from an equity perspective, the evaluation manager/ team will encounter a range of different situations, each requiring a different response as shown in Table 1 below. The table includes several levels of evaluability of equity dimensions, which are to be considered, as well as informa-

³ This section is based on and adapted from *Integrating human rights and gender equality in evaluation*, UNEG, 2011.

tion on the characteristics of interventions and possible approaches to the challenges. In all cases, the evaluation manager/team will have alternative approaches available for addressing evaluability challenges during the evaluation process. An intervention may also present a combination of the characteristics below. In this case, a mixed approach is recommended for dealing with the evaluability challenges.

Table 1. Determining the evaluability of the equity dimensions of the intervention*

Evaluability for equity	Characteristics of the intervention	Possible approaches to address evaluability challenges
High	The intervention theory of change has clearly considered equity issues (e.g. the intervention identified, from the beginning, problems and challenges that affect worst-off groups, inequities and discrimination patterns in the area where it occurs, contextual or systematic violations of rights, etc.)	<ul style="list-style-type: none"> • Make sure that the evaluation ToR takes full advantage of the information already produced by the intervention, and of the participation mechanisms established • Consult stakeholders on whether there are still areas where equity evaluability needs improvement • Address any possible weaknesses and recommend steps to improve evaluability, if necessary. Consult stakeholders on their ideas about how to improve evaluability • If necessary, include methods and tools in the evaluation that can capture new data or strengthen existing data on human rights and equity (e.g. information on new groups of people, changes in the context, etc.) • Use the context (political, institutional, cultural) of the intervention in favor of the evaluation: when it is conducive, build on this support to ensure a highly participatory evaluation
	Equity is clearly reflected in the intervention design (log frame, indicators, activities, M&E systems, reporting mechanisms)	
	The intervention design benefitted from a strong and inclusive stakeholder analysis which included worst-off groups	
	The intervention design benefitted from specific equity analysis	
	Records of implementation and activity reports contain information on how equity was addressed	
	Stakeholders (including worst-off groups) have participated in the various activities of the intervention in an active, meaningful and free manner	
	Monitoring systems have captured equity information (e.g. the situation of different groups, etc.)	

*This table is based on and adapted from "Integrating human rights and gender equality in evaluation", *UNEG, 2011*

Evaluability for equity	Characteristics of the intervention	Possible approaches to address evaluability challenges
High	Data has been collected in a disaggregated manner (e.g. by gender, race, ethnicity, age, etc.) reflecting diversity of stakeholders	<ul style="list-style-type: none"> • Make sure that the equity issues captured in this intervention are also well reflected in the evaluation report • Where a counterfactual, such as a comparison group not affected by the intervention, can be identified; the possibility of using a quasi-experimental design can be considered.
	Progress and results reports for the intervention are equity-focused	
	Context (political, institutional, cultural, etc.) within which the intervention is implemented is conducive to the advancement of human rights and equity	
	It is possible to identify a counterfactual (such as a comparable population that does not benefit from the equity intervention)	
Medium	The intervention's theory of change has considered equity issues to a certain extent, with weaknesses in some areas of the intervention	<ul style="list-style-type: none"> • Understand the reasons for the limitations: are they political, practical, budgetary, time-related, due to limited know-how, etc.? Consult stakeholders and available documentation, which may offer insights on this • Highlight the evaluability limitation in the evaluation ToR. Include, in the evaluation design, tools and methods that make use of the existing data, but that may also help generate new information on equity. Include tools and methods that strengthen stakeholder participation, including worst-off groups • Pay special attention to the stakeholder analysis in the evaluation process, and who should be involved. Make sure that groups who have been left out are considered, and how they could be included at this stage
	Equity has been reflected in the intervention design to some extent (e.g. intended or mentioned, but it is not clearly articulated how to address equity issues in practice; is limited to only a few causes of inequity; addresses numbers without addressing actual changes in inequity; is clear in the narrative but not in the log frame etc.)	
	The intervention design benefitted from a stakeholder analysis, but important worst-off groups have been left out	
	The intervention design benefitted from limited equity analyses	

Medium	Records of implementation and activity reports include limited data on how equity has been addressed	<ul style="list-style-type: none"> • Include in the evaluation process an exercise to strengthen the existing equity analyses • During the evaluation process, seek partners and documents that may have useful information on equity not yet captured by the intervention (e.g. national evaluation/statistics offices, other development agencies, civil society and community organizations, media, academia, etc.) • Build on the context where the intervention is made: if it is conducive to the advancement of equity to a certain extent only, identify key advocates and supporters of the cause and involve them in the evaluation design stage • During the data analysis process, address whether the limitations in the intervention had a negative effect on worst-off groups. Analyze also the negative effect of not being able to substantively assess human rights and equity (e.g. how the lack of this information and data affects the overall evaluation findings, which would basically be incomplete). Consider and consult stakeholders on how this situation could be improved • Include data on equity in the evaluation report, address limitations and provide recommendations for improvement
	Stakeholders, including worst-off groups, have participated in the intervention to a certain extent (e.g. being informed or consulted, but not taking part in decisions; only some groups have been consulted; etc.)	
	Monitoring systems have captured some information on equity	
	Some limited disaggregated data have been collected	
	The context (political, institutional, cultural, etc.) where the intervention is made is conducive, to a certain extent, to the advancement of equity	

Evaluability for equity	Characteristics of the intervention	Possible approaches to address evaluability challenges
Low	The intervention's theory of change failed to consider equity dimensions in its design, implementation and monitoring, or the theory of change does not exist	<ul style="list-style-type: none"> • Reconstruct the missing theory of change and log frame in cooperation with key stakeholders (see section 7 on conducting Equity-focused evaluation under real world constraints) • Understand the reasons for the failure: are they political, practical, budgetary, time-related, due to limited know-how, etc. Consult stakeholders and documentation that may offer insights on this • Highlight the evaluability limitation in the evaluation ToR. Include, in the evaluation design, tools and methods that may help generate information on equity, even if limited. Include tools and methods to enhance stakeholder participation, especially worst-off groups • Pay special attention to the stakeholder analysis in the evaluation process, and who should be involved. Because the equity dimensions have not been considered in the intervention, several important stakeholders will most probably have been left out • Include preparation of equity analyses in the evaluation process
	Stakeholder and/or equity analyses were not conducted adequately or do not exist	
	Data on equity and/or disaggregated data are not available	

Low	Stakeholder (especially worst-off groups) participation in the design, implementation and monitoring processes of the intervention has been minimal	<ul style="list-style-type: none"> • During the evaluation process, seek partners and documents that may have useful information on equity not yet captured by the intervention (e.g. national evaluation/statistics offices, other development agencies, civil society and community organizations, media, academia, etc.) • In spite of the context, try to identify advocates and supporters of equity and involve them at the evaluation design stage • During the data analysis process, pay special attention to the question of whether the intervention has had a negative effect on particular stakeholders, especially worst-off groups. Consider and consult stakeholders on how this situation could be improved • Highlight the challenges of addressing human rights and equity in the evaluation report, also specifically in the evaluation section. Since human rights and equity are a mandate of the UN, which should be considered in every intervention design, provide assertive recommendations for immediate action
	Progress and results reports for the intervention do not address equity issues	
	Context (political, institutional, cultural, etc.) where the intervention takes place is not conducive to the advancement of equity	

3.2 Identifying evaluation stakeholders, including worst-off groups

As already seen above, involvement in the design, planning and implementation of the evaluation of those stakeholders (especially worst-off groups) directly affected by an intervention (be they the implementers or intended beneficiaries), is a fundamental principle of any Equity-focused evaluation.

The degree and level of stakeholder participation in an evaluation process varies and various challenges – institutional, budgetary and time – need to be taken into consideration. However, guaranteeing stakeholder participation strengthens accountability, builds trust and agreement in the evaluation process, generates credibility and can itself contribute to equity building. Evidence also shows that stakeholder participation enhances the use of evaluation conclusions by increasing ownership. The evaluation manager will need to weigh-up the level of stakeholder participation against the benefits and constraints.

Box 1. Determining the degree of stakeholder participation*

The following questions should be considered when deciding the appropriate degree of participation by stakeholders:

1. How can stakeholders, including worst-off groups, be involved in the process with varying degrees of intensity? What will be the implications in terms of effort, timeline and budget?
2. Should all stakeholders be involved together or separately? If involved together, what will be the process for ensuring all perspectives are fairly heard, avoiding bias because some may be more reticent than others for a variety of reasons (power differences, literacy levels, confidence levels, etc.), mediating differences, building agreement, and making decisions where differences cannot be reconciled)?
3. How can the level of participation envisaged by the evaluation process be ensured, even if the reality is that the intervention to be evaluated has had limited participation so far? How can the evaluation generate lessons for the intervention to overcome the participation challenges?
4. 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?

**This box is based on and adapted from Integrating human rights and gender equality in evaluation, UNEG, 2011*

5. Does the evaluation manager have the appropriate level of commitment, understanding, facilitation skills and experience for the level of participation decided upon?
6. Have the gains in credibility of the evaluation results, from a particular level of participation, been considered?
7. Has sufficient consideration been given to participation to ensure the credibility of evaluation results?

As far as possible, stakeholders should be involved in the evaluation from the early stages of the evaluation process, and a *stakeholder analysis* is the most effective tool to help identify who the different groups in an intervention are and why, and how and when they should be included in the evaluation process.

Awareness of the diversity of stakeholders is a critical factor in any process that is sensitive to equity. This means not treating groups, including worst-off groups, as uniform, but understanding and acknowledging that different sub-groups exist and are affected by an intervention in different ways.

A stakeholder analysis is also a helpful tool to address the problem of positive bias in evaluations. Evaluations subject to budget and time constraints interview primarily the direct beneficiaries and implementing agencies for the intervention. Consequently, most of the information received tends to be relatively positive if the intervention is progressing well. Often, however, information is not collected from worst-off groups. However, in Equity-focused evaluations these groups should be involved as much as possible and as appropriate within the local socio-cultural context.

3.3 Identifying the intended use and intended users

UNICEF Evaluation Policy states that the indicator of a successful evaluation function is the strategic use of the evaluation findings. Experience suggests that, while the production of high-quality evaluation reports is a necessary fundamental product, it is not sufficient to ensure the use of evaluation findings.

Any Equity-focused evaluation should determine from the very beginning what the intended use should be and who the intended users are. This is very important to ensure that the purpose and the evaluation questions will be relevant to the information gaps of strategic stakeholders, including worst-off groups. Only when this

is the case will they actually use the evaluation findings in a meaningful manner. However, it is important to be aware that the strategic use of evaluation findings is not only determined by hierarchical positions within an organization or community, but also by real, live, caring human beings. To assure actual use of an evaluation, it is very important to carry out an organizational decision-making analysis to determine:

- a. *who* are the key actors in need of information to solve problems, and,
- b. *who* is likely to use the evaluation findings and to support follow-up actions based on the evaluation's recommendations (Segone, 2006).

This is not meant to imply that only top management should be actively involved in the evaluation process from the start. In fact, very often the key actors are middle managers, officers and stakeholders responsible for developing and implementing the programme in the field. In any case, the personal factor is a key element for guaranteeing the use of evaluation findings. Patton (1997) defines the personal factor as the presence of an identifiable individual or group of people who personally care about the evaluation and its findings. The *personal factor* thus represents the leadership, interest, enthusiasm, determination, commitment and caring of specific individual people. Therefore, when identifying the intended use by intended users, both the organizational structure (leadership and authority) and the personal factor (interest, enthusiasm, and commitment) must be taken into consideration.

Once the intended users, including worst-off groups, are identified, they should be invited – as appropriate given the particular context of each evaluation – to be members of the Steering Committee responsible for: identifying the purpose and scope of the evaluation; the key evaluation questions; approval of the Term of Reference, including the evaluation framework and methods, and the final report; and, most importantly, for leading the dissemination, communication and use of the findings.

3.4 Identifying potential challenges in promoting and implementing Equity-focused evaluations

Various challenges can be faced when promoting and implementing Equity-focused evaluations. A list of some potential challenges is given below. This list should help the evaluation manager and stakeholders to identify challenges early on in the evaluation process and to develop relevant strategies to overcome them.

Potential challenges in promoting Equity-focused evaluations are as follow:

Reluctance to accept disaggregated indicators, which can show country performance in a poor light. For example, many countries have been using the monetary definition of poverty, estimating the proportion of the population below the poverty line. Based on this indicator many countries have made significant progress, and therefore they can report steady progress to the international community and to national policy makers and public opinion. Similarly, many countries have been making progress towards the MDGs and this has improved their world-ranking on the UNDP Human Development Index. Shifting to an equity-focused analysis will frequently identify additional groups who can be defined as vulnerable through one or more indicators, and this can show countries in a less favorable light. For this reason some countries may be reluctant to adopt the equity focus. This can be particularly true in countries with upcoming elections, where the government wishes to be able to report positive progress in addressing poverty and social problems.

Political and social resistance to addressing the causes of exclusion and vulnerability. Many of the causes of vulnerability and social exclusion are social or political. Resources may not be channeled to certain areas because they are from different ethnic or tribal groups that do not support the party in power; landowners may not wish to recognize the property rights of the landless and groups occupying land claimed to be private; religious minorities may be in conflict with dominant religious groups; the continuation of the caste system is justified on religious grounds. Consequently, there may be active or at least passive opposition to addressing the root causes of exclusion.

Resistance to empowerment of worst-off groups. Many political systems are based on patronage whereby parties and politicians chan-

nel resources in return for political support. Empowering worst-off groups and involving them in decisions over resource allocation would erode the basis of political power and influence. Consequently political parties may be opposed to empowerment or even to the use of more rational targeting mechanisms to allocate resources.

Lack of interest/incentives and reluctance to invest resources in the worst-off groups. Some countries have no incentive to focus on equity because reaching worst-off groups is not a national priority. This may be the case particularly in countries with a very large poor population with limited access to services. In addition, politicians, and often voters, are unwilling to invest resources in programmes for worst-off groups as this may mean reducing services to the politically more influential groups.

Governance. More effective targeting, and delivery of services to worst-off groups is dependent on the capacity of public service agencies to manage programmes and deliver services effectively. Most services are delivered at the local level, so effective decentralization is often a requirement. In many countries governance is an issue and there is only limited decentralization of authority and resources. Consequently, government agencies do not have the capacity to design and implement programmes targeted at worst-off groups. The situation is even more difficult in countries where corruption is an issue, because worst-off groups, by definition, are those with the least voice and least ability to defend their rights.

The legal status of worst-off groups. Delivering services to vulnerable groups is complicated by the fact that many of them lack full legal documentation or full property rights to their home or land. Many live on land for which they hold no formal occupation rights, and there is strong pressure from former occupiers or property developers not to provide services as this would implicitly recognize their right to occupy the land. Under these circumstances governments are often unwilling to develop services for these groups and hence there is little interest in conducting Equity-focused evaluations.

Potential challenges in implementing Equity-focused evaluations are as follows:

Methodological challenges in the evaluation of complex interventions. Equity-focused interventions are more complex than “conventional” ones. Equity-focused evaluations, especially at policy-level, must therefore use innovative approaches to evaluate complex

interventions. However, so far, the evaluation literature only provides emerging guidance on how to evaluate outcomes and impacts for these kinds of complex interventions.

Lack of disaggregated data or data collection capacity, and reluctance to change existing methodologies. Equity-focused evaluation requires more detailed data and often larger sample sizes, and this may not exist in many countries. Many countries may not have the financial, logistical or technical resources to collect and process these additional data. While there are many creative ways to introduce an equity focus while operating under budget and time constraints (for example using mixed-method approaches), countries may have neither the expertise nor the incentives to do this.

Additional cost and complexity. The required budget may not be available to conduct more expensive equity-focused evaluations, even if agencies would be willing to carry-out these evaluations.

The need to base the programme and the evaluation on a programme theory of change. The more in-depth analysis required for Equity-focused evaluation requires that the evaluation be based on a programme theory of change, so that hypotheses can be developed and tested about behavioral, cultural and other factors affecting implementation; and about how implementation and outcomes are affected by contextual factors. Ideally the programme's theory of change will have been developed as part of programme design and then adopted by the evaluation. Where this has not been done the theory of change can be reconstructed retrospectively by the evaluation team. However, pressure to start the programme and to begin to deliver services means that time and resources are often not available to develop this model. This makes it much more difficult to develop a rigorous evaluation design.

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

In addition to the above issues and challenges, there are **further challenges specific to Equity-focused evaluations in a humanitarian setting**. By their very nature, humanitarian crises often

create a category of those worst-off in society: those most affected by the crisis. Even within the affected population, however, the effects of the emergency can vary widely, whereby some parts of the population are affected far more than others, by virtue of geography (those directly in the path of the conflict or natural disaster); socioeconomics (those less equipped with the necessary resources to rebuild and return to normality); or social location (those with less of a voice in society and therefore in the response, or those on the losing side of a conflict, and those whose mobility is further challenged by disability). Children often bear the brunt of emergencies, owing not only to their injury or illness, but also through the loss of, injury to, or separation from the protective influences of parents and family, and the destruction and displacement of schools and communities. Thus, it is difficult to conduct Equity-focused evaluations in emergency settings for these reasons, as well as the inherent technical difficulty in these settings – i.e., accessing affected areas – which can add considerable cost. It can be particularly difficult to systematically determine who are the “worst-off among the worst-off” – and to reach them. This can be particularly difficult when structural sources of inequity are at play. Such structural factors can limit government support for deliberately accessing these sub-populations with a view to assessing their experience, or for seeking their participation in the evaluation, especially in situations where the government is party to the conflict at hand.

Further challenges are of a more strategic nature. These include the frequent perception among operational colleagues and management that evaluation is “getting in the way” of their response work – not least when the evaluation might uncover shortcomings in reaching the worst-off, and might reveal that the response has reinforced or even worsened inequity, rather than redressing it. Thus, it is important that Equity-focused humanitarian evaluations, even those with learning as a primary goal, are properly anchored in the language of accountability to the affected populations, including the most vulnerable – children and women.

SECTION 4: PREPARING THE EVALUATION TERMS OF REFERENCE⁴

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

The evaluation manager, together with the Steering Committee, will have the greatest influence in shaping the evaluation planning stage: deciding the purpose, scope and focus of the evaluation, including developing the Terms of Reference (ToR). It is therefore important that the evaluation manager has a good understanding of Equity-focused evaluation. Otherwise, assistance, especially in planning and developing the ToR for the evaluation, should be sought.

4.1 Defining the scope and purpose of the evaluation

As a first step, the Steering Committee should clearly define the purpose of the evaluation, including why the evaluation is needed at this stage, who needs the information, what information is needed, and how the information will be used. A clear explanation of the evaluation objectives and scope, including main evaluation questions, should be developed, paying attention to keeping both purpose and scope focused. This will help the evaluation manager to lead the process of developing the Terms of Reference, including framing the evaluation questions.

4.2 Framing the evaluation questions

The evaluation questions, together with the purpose and scope, are the central part of the ToR. They will inform the decision on what methodology the evaluation should use.

As with the other tools in this document, these examples of questions need to be considered in context, and adapted to the specific reality of the intervention to be evaluated. The questions must derive from the theory of change for the intervention, which is specific to the intervention, and it should be noted that there will

4 This chapter is based on and adapted from “Integrating human rights and gender equality in evaluation”, UNEG, 2011

always be issues that cannot be pre-empted in guidance material. As already noted in the evaluability section, it may be the case that an intervention does not have an explicit theory of change. In this case, an evaluation can also reconstruct the implicit theory of change for an intervention (see section 7 on conducting Equity-focused evaluation in real-world settings). The questions in Table 2 provide the starting point for a more profound investigation. Probing for further details, underlying reasons, alternative scenarios etc., is critical to answering the questions and these qualitative refinements will help evaluators reach the more complex answers.

Evaluation criteria provide an overarching framework for an evaluation and define the evaluation questions. The UN commonly uses and adapts the evaluation criteria from the Organization for Economic Cooperation and Development's Development Assistance Committee (OECD-DAC) to evaluate its interventions. These are relevance, effectiveness, efficiency, impact and sustainability⁵. Many organizations add their own additional criteria such as: gender equality, knowledge management and developing an effective Monitoring and Evaluation system. Additional criteria, such as the Active Learning Network for Accountability and Performance (ALNAP) humanitarian criteria, are also commonly used.

However, the mainstream definitions of the OECD-DAC criteria, as well as the ALNAP ones, are neutral in terms of equity dimensions. Table 2 provides some guidance on how to integrate equity dimensions into the OECD-DAC and ALNAP evaluation criteria, when proposing potential Equity-focused evaluation questions.

5 See DAC Criteria for Evaluating Development Assistance (OECD-DAC 2010)

Table 2: Evaluation criteria and potential questions for Equity-focused evaluations

DAC criteria adapted for Equity-focused evaluations	Potential Equity-focused evaluation questions
<p>Relevance: The extent to which the expected results of the intervention address the rights and needs of worst-off groups, reduce inequities, and are consistent with equity-focused development priorities at global, national or local level.</p>	<ul style="list-style-type: none"> • What is the value of the intervention in relation to the needs of the worst-off groups, reduction of inequities between the best-off and the worst-off groups, equity-focused national priorities, and national and international partners' equity-focused policies? • What is the value of the intervention in relation to global references such as human rights, humanitarian law and humanitarian principles? • What is the relevance in relation to the equity approach, as well as foundation strategies such as the Human Rights-based Approach to Programming and Gender Mainstreaming, the Core Commitments for Children in Humanitarian Action (CCCs), and, in the case of UNICEF-supported interventions, Medium Term Strategic Plan (MTSP)? • What does the literature and current experience suggest about the appropriateness of the current or proposed strategy? If successfully implemented would this strategy be likely to address the key issues affecting the worst-off groups?
<p>Impact: Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended, for the worst-off groups as well as inequities between best-off and worst-off groups.</p>	<ul style="list-style-type: none"> • What are the results of the intervention – intended and unintended, positive and negative – including the social, economic, and environmental effects on the worst-off groups? • How do the results affect the rights and responsibilities of worst-off individuals, communities and institutions? • To what extent have results contributed to decreased inequities between the best-off and the worst-off groups?

<i>DAC criteria adapted for Equity-focused evaluations</i>	<i>Potential Equity-focused evaluation questions</i>
<p>Effectiveness: The extent to which the intervention's equity-focused results were achieved, or are expected to be achieved.</p>	<ul style="list-style-type: none"> • Is the intervention achieving satisfactory results in relation to stated equity-focused objectives? • How does quality of public systems targeting worst-off groups compare with the quality of conventional public systems? • Were contextual factors (political, social, economic, cultural) taken into account in the design/implementation of the intervention? • Are public and private service delivery systems reaching the worst-off groups? <ul style="list-style-type: none"> – What are the main constraints on supply? – What are the main constraints on demand? – Which programmes are most and least effective? – What factors explain success?
<p>Efficiency: A measure of how economically resources/inputs (funds, expertise, time, equipment, etc.) are converted to equitable results.</p>	<ul style="list-style-type: none"> • Does the programme use resources in the most economical manner to achieve expected equity-focused results? • Are any other economical alternatives feasible? • How cost-effective are the public systems for reaching worst-off groups? <ul style="list-style-type: none"> – How do costs for reaching worst-off groups compare with average public services costs? – How do costs for reaching worst-off groups compare with alternative systems to deliver services to worst-off groups?

<p>Sustainability: The continuation of benefits to the worst-off groups after major development assistance has been completed. Sustainability looks to the probability of continued long-term benefits for the worst-off groups.</p>	<ul style="list-style-type: none"> • Is the intervention and its impact on the worst-off groups likely to continue when external support is withdrawn? • Are inequities between best-off and worst-off groups likely to increase, remain stable, or decrease when external support is withdrawn? • Will the strategy be more widely replicated or adapted? Is it likely to be scaled-up (“go to scale”)?
<p>Additional criteria in humanitarian evaluations*</p>	<p>Potential Equity-focused evaluation questions</p>
<p>Coverage: The need to reach population groups facing life-threatening suffering, wherever they are marginalized geographically, socio-economically, or by virtue of their social location.</p>	<ul style="list-style-type: none"> • How closely have humanitarian principles of humanity, impartiality and neutrality – all having direct linkages to equity – been followed in preparing for, and responding to, the emergency at hand? • What proportion of the affected population has been reached overall? Did the intervention reach the worst-off groups to the same extent that it reached other groups? • What key barriers have hindered all of the affected population being reached, especially those most affected by the emergency (e.g., geographic remoteness, security situation, etc.)? How successfully have the barriers to reaching them been identified and overcome? • Within the overall affected population, have any groups been affected worst? How successfully have the barriers to reaching them been identified and overcome? • In what way, if any, has the emergency at hand affected the equity profile in the affected population – e.g., to what extent have pre-existing sources of inequity been exacerbated, to what extent have power relations shifted, and so on? How adequately have needs assessments detected these shifts, and how sufficiently have they been taken into account in the response?

* *Criteria definitions adapted from those articulated in: Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies. London: Overseas Development Institute, March 2006.*

Additional criteria in humanitarian evaluations*	Potential Equity-focused evaluation questions
<p>Connectedness: The need to ensure that activities of a short-term emergency nature are carried out in a context that takes longer-term and inter-connected problems into account – in particular the need to “build back better” in a way that serves to redress rather than to reinforce or worsen inequity, and to address the equity-rooted sources of conflict and natural disasters.</p>	<ul style="list-style-type: none"> • How well prepared were the affected country and regional offices for the shift to emergency operations – e.g., through well-crafted Emergency Preparedness & Response Plan (EPRPs) and Business Continuity Plan (BCPs), simulation exercises, and so on? • How quickly were early recovery activities integrated into the response? How well tailored were these to ensure that the worst-off were provided with appropriate support, proportional to their specific needs, to bring their chance of recovery on a par with that of the rest of the affected population? • To what extent were short-term emergency activities carried out with a view to long-term recovery – and in particular the need to “build back better” in a way that redresses key sources of inequity? How well integrated into longer-term planning processes were disaster risk reduction and peace-building activities, so as to redress the equity-related sources of conflict and natural disasters?
<p>Coherence: The need to assess security, development, trade and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations.</p>	<ul style="list-style-type: none"> • What is the level of coherence around equity in the guiding policies of different humanitarian actors? Are equity considerations explicitly taken into account in these policies? • How effectively has UNICEF leveraged its position to advocate for a more equitable policy environment, e.g., by anticipating and addressing structural sources of heightened vulnerability before the emergency strikes, securing humanitarian space at its onset – for maximum coverage, effectiveness, and so on? • How effective has the coordination effort been, either through the cluster approach (in terms of intra and inter-cluster coordination) or alternative modes of coordination, to ensure that the response reaches the affected population, particularly the worst-off groups? • How effectively have those involved in the humanitarian response collaborated with non-traditional actors (e.g., the corporate sector, military, and so on) to help pave the way to reach the worst-off without betraying humanitarian principles?

*Criteria definitions adapted from those articulated in: Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies. London: Overseas Development Institute, March 2006.

4.3 Selecting a technically-strong and culturally-sensitive evaluation team

Selecting a strong team to conduct an Equity-focused evaluation is a key step in a successful evaluation process. A good team must have an appropriate mix of skills and perspectives. The team leader is responsible for organizing the work distribution, and for making sure that all team members contribute meaningfully. Insofar as possible, the following attributes and capacities should be included in the team:

- Balance in diversity, including gender, ethnicity, etc.
- Evaluators from the worst-off groups (or at least very familiar with the contexts and situation of worst-off groups).
- Local and/or international evaluators.
- Evaluation knowledge and experience (quantitative and qualitative methods).
- Content/sectoral knowledge and experience.
- Commitment to human rights and equity, and knowledge and experience in evaluating human rights and pro-equity interventions.
- Understanding and application of UNICEF mandates on human rights and equity.
- Experience in, and knowledge of, participatory approaches and methods.
- Research and interpersonal skills, including cultural competence.
- Knowledge of regional/country/local context and language.

In putting together an evaluation team, one important aspect needs to be taken into consideration. It is common to see teams reproducing the same imbalances and patterns that exist in real life. What makes a good evaluation team for addressing equity is not only the skills and competences held collectively by its members, but also the dynamics of the interactions between them. Team members must demonstrate their capacity to appreciate and include each other's expertise and perspectives. The evaluation manager must ensure that appropriate weight is given to the equity dimensions both through the team selection and attention to the dynamics and relations among team members. Working with a multidisciplinary

team will most often be the ideal approach for dealing with the complexities of evaluating an intervention.

When team members come from diverse backgrounds it is important to invest time and resources in team-building prior to the start of the evaluation. Mutual respect for the personal and professional background of other team members and for understanding their different approaches to research and evaluation are critical to a successful Equity-focused evaluation.

SECTION 5: DESIGNING THE EVALUATION

This section highlights the importance of using appropriate methods for an Equity-focused evaluation, to ensure that the equity dimensions of the intervention will be identified and analyzed during the evaluation process.

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

5.1 Selecting the appropriate evaluation framework

Below are two frameworks, and a number of designs and tools, which can be taken into consideration when planning an Equity-focused evaluation. Many other frameworks, designs and tools relevant and suitable for Equity-focused evaluations exist. The final decision on what framework, design and tools should be used has to be based on the purpose and scope of the evaluation, the evaluation questions, and also the nature and the context of the intervention to be evaluated.

A. Theory-based Equity-focused evaluation

While the programme's theory of change is an important component of most programme evaluations, a well-articulated theory of change is particularly critical for Equity-focused evaluations. Equity interventions achieve their objectives through the promotion of behavioral changes that cannot be defined and assessed through conventional pre-test/post-test comparison group designs comparing a set of indicators before and after the intervention. The process of implementation, and the context within which implementation takes place, have a significant impact on the accessibility of the health, education and child protection public systems for worst-off groups. It is also important to understand how effectively public policies and service delivery systems have been able to adapt to the special challenges of reaching worst-off groups. For all of these reasons it is important to base the evaluation on a theory of change that can describe and assess the complex reality within which Equity-focused interventions operate.

A well-articulated programme theory of change can:

- Define the nature of the problem the policy or programme is intended to address.
- Incorporate lessons from the literature and experiences with similar programmes.
- Identify the causes of the problem being addressed, and the proposed solutions.
- Explain why the programme is needed.
- Identify the intended outcomes and impacts.
- Present a step-by-step description of how outcomes are to be achieved.
- Define the key assumptions on which the programme design is based.
- Identify the key hypotheses to be tested.
- Identify the contextual factors likely to affect implementation and outcomes.
- Identify the main risks and reasons why the programme may not achieve its objectives.

The programme theory is also a valuable tool in the interpretation of the evaluation findings. If intended outcomes are not achieved, the programme theory can help trace-back through the steps of the results chain to identify where actual implementation experience deviated from the original plan. It also provides a framework for identifying unanticipated outcomes (both positive and negative). If implementation experience conforms reasonably closely to the design, and if outcomes are achieved as planned, this provides prima facie evidence to attribute the changes to the results of the programme. However, it is possible that there are other plausible explanations for the changes, so a well-designed programme theory should be able to define and test rival hypotheses. The theory must be defined sufficiently precisely that it can be “disproved”. One of the major criticisms of many programme theories is that they are stated in such a general and vague way that they can never be proved wrong. To disprove requires:

- that the theory includes a time-line over which outcomes and impacts are to be achieved;

- measurable indicators of outputs, outcomes and impacts;
- measurable indicators of contextual factors, and a clear definition of how their effect on policy implementation and outcomes can be analyzed.

Ideally the programme's theory of change will be developed during the policy design. However, it is often the case that the theory of change was not developed so the evaluation team must work with stakeholders to "reconstruct" the implicit theory on which the policy is based (see section 7 on Real World evaluation). Ideally the evaluation team will be involved at a sufficiently early stage of the design to be able to assist in the development of the programme's theory of change, so as to ensure that it provides sufficient detail for the evaluation.

Basic components of a programme theory of change

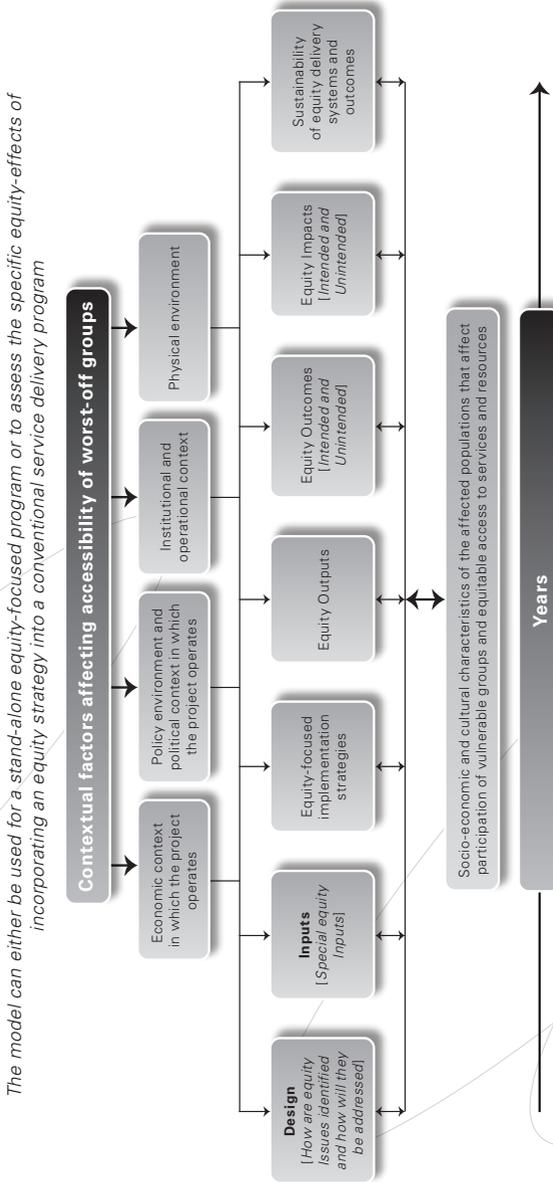
Programme theories of change are often represented graphically through a logic model. Figure 1 presents a typical logic model describing an equity-focused intervention designed to ensure that services and benefits of a programme are accessible to specific worst-off groups. The model can be used either to describe a stand-alone programme targeted at worst-off groups (for example female sexual partners of injecting drug users), or to describe equity-focused strategies that are integrated into a universal programme. An example of the latter would be a programme designed to increase overall school enrolment through separate toilets for boys and girls, renovated buildings, new school textbooks and teacher training programmes. A special scholarship programme and transport vouchers might be targeted specifically at girls from low-income households to provide a further incentive for them to enroll. In this case the evaluation would assess the overall impacts of the programme on school enrolment as well as the effectiveness of the scholarships and transport vouchers on increased enrolment for low-income girls. If resources permit the evaluation might use the programme theory as a framework to compare enrolment rates for low-income girls in schools that only offered the general improvement programmes with those that also included the targeted programmes. This would permit an analysis of the value-added of the targeted programmes.

The model includes two main components:

- The seven stages of the project cycle (design, inputs, implementation, outputs, outcomes, impact and sustainability) – defining the special equity-focused elements at each stage.

- The contextual factors (political, economic, institutional, natural environment and socio-cultural characteristics of the affected populations) that can affect the implementation and outcomes of equity-focused interventions.

Figure 1: Using a logic model to represent a program theory of an equity-focused intervention



Refinements to the basic logic model

There are a number of refinements that can be incorporated in the basic logic model that are important for the description and evaluation of equity-focused interventions:

- *The contextual framework*: analysis of the economic, political, socio-cultural, environmental, legal, institutional and other factors, that affect how programmes are implemented and how they achieve their outcomes. All of these factors can constrain the effective implementation of equity-focused interventions. In cases where there is little social or political support for the integration of worst-off groups, many of these factors can present major challenges. While contextual factors are often analyzed descriptively, it is also possible to incorporate these variables into the statistical analysis by converting them into dummy variables.
- *Process analysis*: examining how the programme is actually implemented, how this compares with the intended design, and the effects of any deviations from the design, and how deviations affect the accessibility of the programme for different sectors of the target population.
- *Results chain analysis (also called outcomes chain)*: a step by step explanation of how the programme is expected to operate and how it will achieves its objectives.
- *Trajectory analysis*: defining the time horizons over which different outcomes are expected to be achieved.

Box 2. Useful references for understanding programme theory

- Funnell and Rogers (2011). *Purposive programme theory: effective use of theories of change and logic models*. Jossey-Bass Publications
- Patton (2011). *Developmental Evaluation: applying complexity concepts to enhance innovation and use*. Guilford Press.
- Bamberger, Rugh and Mabry (2006). *Real World Evaluation*. Chapter 9 “Applications of programme theory in real world evaluation” Sage Publications

B. The bottleneck analysis framework

Bottleneck supply and demand analysis has been used successfully to evaluate service delivery systems, especially in health systems. It provides a framework for the description and analysis of the major

factors affecting the access of worst-off groups to public services, and it has the potential to be an integrated tool that can identify the strengths and weaknesses of different service delivery systems. However, it is important to note that this framework has important limitations when evaluating interventions dealing with acts of commission rather than omission, notably in the field of child protection and violence against children and women.

The framework has four components (see Figure 2):

Use of services by worst-off groups

Defining the worst-off groups to be targeted by the intervention. A first step is to identify the worst-off groups intended to benefit from the intervention. The groups can be defined geographically (for example, living in a particular district or in all rural areas) as well as by the nature of the inequity (gender, ethnicity, etc.).

Assessing the adequacy of service utilization by worst-off groups. The following measures should be combined, as appropriate, to assess effectiveness in delivering quality services to the target worst-off groups. Performance indicators include:

- The proportion of each worst-off group who utilize the service.
- The adequacy level for utilization of each service.⁶
- A comparison of the proportion of the total population utilizing the service with the proportion of the worst-off group who utilize it.
- A comparison of the adequacy of utilization by worst-off and by other groups.

Assessing sustainability. Many interventions operate well whilst donor agencies are actively involved or whilst special programme funding is available, but the quality or volume of services often declines when these special incentives end. It is therefore important to continue to monitor programme operations over time in order to assess long-term sustainability.

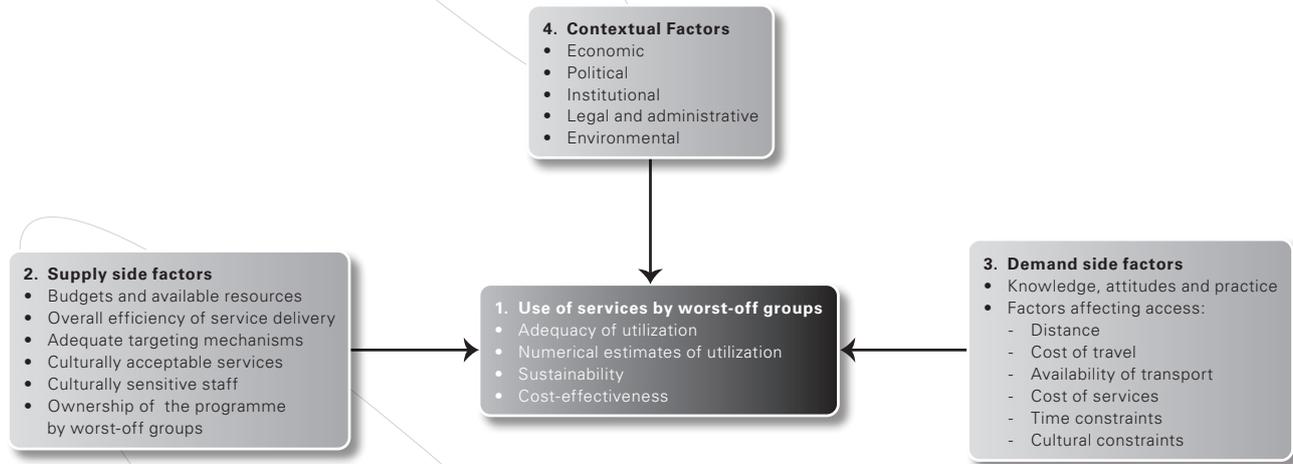
6 Examples of indicators of adequacy of utilization include: does the pregnant mother sleep under the bed-net provided through the programme? Does the under-nourished child receive the entire intended nutritional supplement or is it shared with siblings? Do women and children receive the medical services free of charge (as intended) or do they have to pay out-of-the pocket to health centre staff?

Identifying different scenarios for access to worst-off populations. The indicators can be used to identify different scenarios, each of which has different policy and operational implications. For example:

- A programme reaching a high proportion of the total population but a low proportion of the worst-off groups. This indicates that there are some specific problems in reaching the worst-off groups.
- A programme reaching only a low proportion of both populations. This suggests that the overall programme performance needs to be improved before greater access to the worst-off groups can be expected.
- The adequacy of utilization by worst-off groups is lower than for other groups. This suggests that there are some specific delivery issues to be addressed.
- Only a small proportion of the worst-off group use the service but the adequacy of utilization is high for this group. This suggests the programme design can potentially benefit the worst-off group but that there are problems in ensuring access.

Cost-effectiveness analysis. Cost will often be a critical factor when budgets are limited or when the equity interventions do not enjoy broad political support. Consequently the analysis of costs, and how they can be reduced, will often be a critical determinant of the success and sustainability of the intervention.

Figure 2: Bottleneck supply and demand framework: factors affecting the use of services by worst-off groups



Supply side factors

The following supply-side factors are assessed:

- Budgets and resources such as staff, buildings, transport, school supplies.
- Overall efficiency of the service organization and delivery.
- Adequate targeting mechanisms. How well does the programme identify the worst-off groups? How adequate are the administrative and other mechanisms for reaching them?
- Culturally acceptable services. Are the services designed in a way that is acceptable to the worst-off groups? For example, many indigenous cultures do not accept the way that western medicine is delivered. Men may not allow their wives or daughters to visit health centres.
- Culturally sensitive staff. Are staffs familiar with the characteristics of the worst-off groups and do they understand the special issues involved in working with these groups? Do they have a positive attitude to working with these groups? Are there staff members who speak the local languages?
- Do worst-off groups have “ownership” of the programme? Were they consulted on how it was designed? Are they involved in management, monitoring and evaluation?

Please note that supply-side issues will be different for special stand-alone programmes targeted exclusively at worst-off groups and for universal service delivery systems adapted to reach worst-off groups.

Demand side factors

The achievement of equity outcomes usually involves processes of behavioral change for different actors. Even when there is a demand for services, and when they are designed in a culturally appropriate way, there are a number of logistical and cultural factors affecting access:

- Distance to the service.
- Time, cost and availability of transport.
- Acceptability of the transport to worst-off groups and their being allowed to use it.
- Costs of services.

- Time constraints.
- Cultural constraints.

Contextual factors

The accessibility of services to worst-off groups can be affected by a wide range of local, regional and national contextual factors including:

- *Political factors.* The attitude of different political groups to providing services to the worst-off (e.g. are worst-off groups considered a security threat; a nuisance; a financial burden; a potential base of political support; a moral obligation; etc.)
- *Economic factors.* The state of the local and national economy can affect the availability of resources. When the local economy is growing, worst-off families may have more incentive to send their children (particularly girls) to school if they are more likely to find employment when they leave.
- *Institutional and organizational factors.* How well do different agencies work together to coordinate services?
- *Legal and administrative.* Do worst-off groups have all of the documents required to access services? Are they registered with the appropriate agencies? Are there legal constraints on providing services to, for example, families who do not own the title of the land they farm or on which they live? Can the ministry of education build schools on land without title?
- *Environmental.* How is programme delivery or sustainability affected by environmental factors such as soil erosion or salinization; flooding; deforestation; water contamination; air quality; or the proximity of urban waste?

5.2 Selecting the appropriate evaluation design

Irrespective of the size and nature of the intervention, an evaluation design which applies a mixed-method approach will usually be the most appropriate to generate an accurate and comprehensive picture of how equity is integrated into an intervention. Mixing qualitative and quantitative approaches, while ensuring the inclusion of different stakeholders (including the worst off groups), will offer a wide variety of perspectives and a more reliable picture of reality.

A. Mixed-methods designs

Mixed-method designs combine the strengths of quantitative (QUANT) methods (permitting unbiased generalizations to the total population; precise estimates of the distribution of sample characteristics and breakdown into sub-groups; and testing for statistically significant differences between groups) with the ability of qualitative (QUAL) methods to describe in depth the lived-through experiences of individual subjects, groups or communities. QUAL methods can also examine complex relationships and explain how programmes and participants are affected by the context in which the programme operates.

These benefits are particularly important for Equity-focused evaluations where it is necessary to obtain QUANT estimates of the numbers and distribution of each type of inequity but where it is equally important to be able to conduct QUAL analysis to understand the lived-through experience of worst-off groups and the mechanisms and processes of exclusion to which they are subjected. QUAL analysis is also important to assess factors affecting demand for services and to observe the social, cultural and psychological barriers to participation.

One of the key strengths of mixed-methods is that the sample design permits the selection of cases or small samples for the in-depth analysis that are selected, so it is possible to make statistically representative generalizations to the wider populations from which the cases are selected. This is critical because typically when QUANT researchers commission case studies to illustrate and help understand the characteristics of the sample populations, little attention is given to how the cases are selected and how representative they are. Very often the case studies, because of their ability to dig more deeply, will uncover issues or weaknesses in service delivery (such as sexual harassment, lack of sensitivity to different ethnic groups, lack of respect shown to poorer and less educated groups, or corruption). When these findings are reported it is difficult to know how representative they are, and consequently it is easy for agencies to dismiss negative findings as not being typical. Mixed-method samples can ensure that the case studies are selected in a representative manner.

However, it is important to highlight the fact that mixed-method designs involve much more than commissioning a few case studies or focus groups to complement a quantitative sample survey. It is an integrated evaluation approach, applying its own unique methods at each stage of the evaluation.

Mixed-methods for *data collection* combine quantitative methods such as surveys, aptitude and behavioral tests, and anthropometric measures with the QUAL data collection methods such as observation, in-depth interviews and the analysis of artifacts⁷. QUAL methods can also be used for process analysis (observing how the project is actually implemented and how these processes affect the participation of different groups within the vulnerable population). The following are examples of how mixed-methods combine different QUANT and QUAL data collection methods:

- Combining QUANT questionnaires with QUAL in-depth follow-up interviews or focus groups.
- Combining QUANT observation methods with QUAL in-depth follow-up interviews.
- Combining QUANT unobtrusive measures with QUAL in-depth interviews or case studies.

In addition, mixed-methods can combine QUANT and QUAL *data analysis* methods in the following ways:

- Parallel QUANT and QUAL data analysis: QUAL and QUANT data are analyzed separately using conventional analysis methods.
- Conversion of QUAL data into a numerical format or vice versa.⁸
- Sequential analysis: QUANT analysis followed by QUAL analysis or vice versa.
- Multi-level analysis.
- Fully integrated mixed-method analysis.

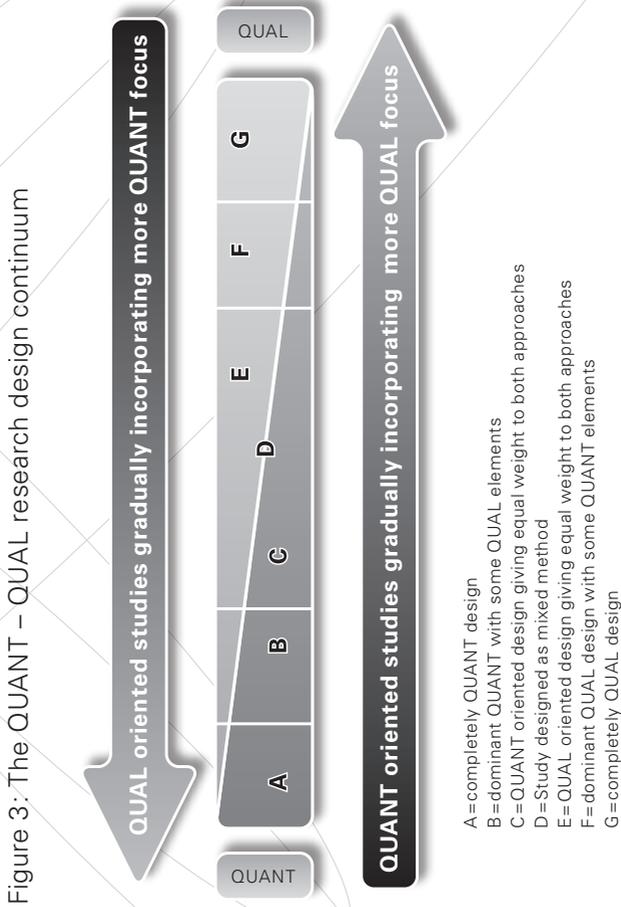
Types of mixed-method designs

Most mixed-method designs are used by researchers who have either a QUANT orientation and recognize the need to build-in a QUAL component, or researchers with a QUAL orientation who recognize the need to build in a QUANT component. Very few mixed-method designs give equal weight to both approaches. Mixed-method designs (Figure 3) can be considered as a continuum with completely QUANT design at one end and completely QUAL at the

7 Examples of artifacts are: photographs and religious symbols in houses, clothing styles, posters and graffiti and different kinds of written documents.

8 An example of conversion of a QUAL indicator into a QUANT variable would be when a contextual analysis describes the status of the local economy. This may be described in words. This can be converted into a dummy variable (Local economy) where: Economy is growing = 1; economy is not growing = 0.

other. In between are designs that are mainly QUANT with a small QUAL component, designs that are completely integrated with equal weight given to QUANT and QUAL, to designs that are mainly QUAL with only a small QUANT component.



There are three main kinds of mixed-method 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 are dominant. Figure 4 gives an example of a sequential mixed-method evaluation of the adoption of new seed varieties by different types of farmer. The evaluation begins with a QUANT survey to construct a typology of farmers and this is followed by QUAL data collection (observation, in-depth interviews) and the preparation of case studies. The analysis is conducted qualitatively. This would be classified as a sequential mixed-method design where the QUAL approach is dominant.
- *Parallel*: The QUANT and QUAL components are conducted at the same time. Figure 5, which illustrates a multi-level evaluation of a school feeding programme might also include some parallel components. For example, QUANT observation checklists of student behavior in classrooms might be applied at the same time as QUAL in-depth interviews are being conducted with teachers.
- *Multi-level*: The evaluation is conducted on various levels at the same time, as illustrated by the multi-level evaluation of the effects of a school feeding programme on school enrolment and attendance (Figure 5). The evaluation is conducted at the level of the school district, the school, classrooms and teachers, students and families. At each level both QUANT and QUAL methods of data collection are used. Multi-level designs are particularly useful for studying the delivery of public services such as education, health, and agricultural extension, where it is necessary to study both how the programme operates at each level and also the interactions between levels.

Figure 4: Sequential QUAL dominant mixed methods design

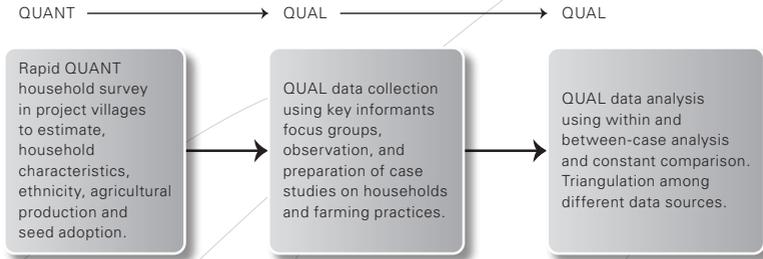
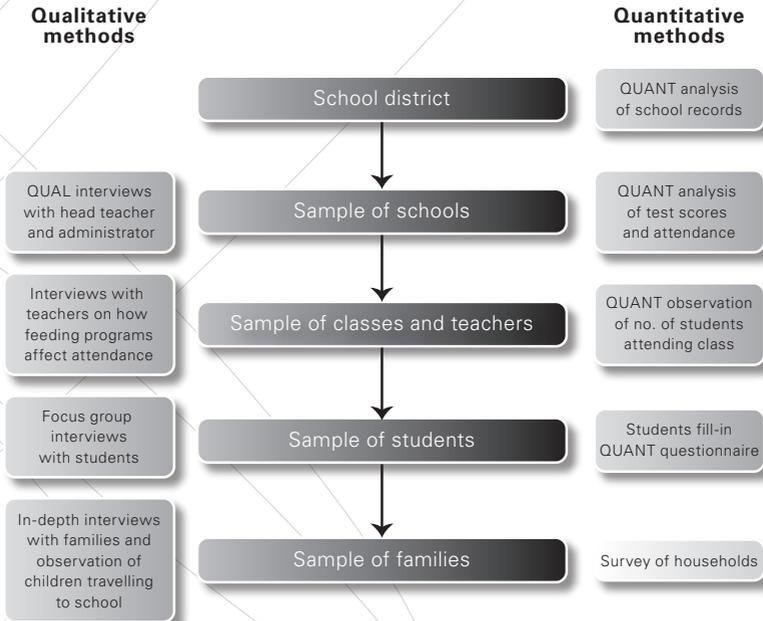


Figure 5: Parallel, multi-level mixed methods design



Box 3 presents an example of a mixed-method evaluation design used to evaluate the equity outcomes of a UNICEF-supported education programme in Timor l'Este. The case shows how a creative mixed method design can produce useful and credible evaluation results in a context where access to quantitative data was very limited. It also illustrates how triangulation was used to strengthen the reliability of the data collected from focus groups, observation and in-depth interviews on the findings; and the validity of the findings.

Box 3. Using a mixed-method design to evaluate the equity outcomes of the UNICEF education programme in Timor l'Este

Timor l'Este is one of the poorest countries in the world, and poverty combined with a history of conflict has seriously affected the quality and accessibility of the school system. One of the goals of the UNICEF supported education programme (2003-09) was to increase the accessibility of the education system for all sectors of society, specifically targeting vulnerable groups including children suffering from HIV/AIDS and children not attending school at all. The evaluation, although not at that time called an Equity-focused evaluation, specifically addressed the effectiveness of the programme in increasing access for worst-off groups.

The original evaluation design planned to combine collection and analysis of quantitative data, from both surveys and secondary data, with more in depth quality methods. However, the paucity of quantitative data meant that greater reliance had to be placed on qualitative data sources. The principal data collection methods were a sample of focus groups selected to be representative of districts and sub-districts throughout the country, combined with structured interviews and direct observation of a sample of schools and how they were operating. The selected districts were chosen to represent the programme diversity in terms of levels of programme saturation ("dosage"), as well as geographic, linguistic and religious variation. Separate focus groups were conducted with pupils, teachers, school administrators, community members, youth and district education officials. The primary data was complemented by an analysis of the extensive secondary data available from project records and other sources. Secondary data was used as an independent source to triangulate with primary survey data in order to test for consistency. Secondary data also expanded the scope of the evaluation as it was possible to compile information on the first phase of the project where, due to the passage of time, it was difficult to locate and identify pupils and the other groups covered by the focus groups.

Triangulation: a powerful tool for assessing validity and for deepening understanding

Triangulation is a very powerful element of the mixed-method approach. It involves using two or more independent sources to assess the validity of data that has been collected and to obtain

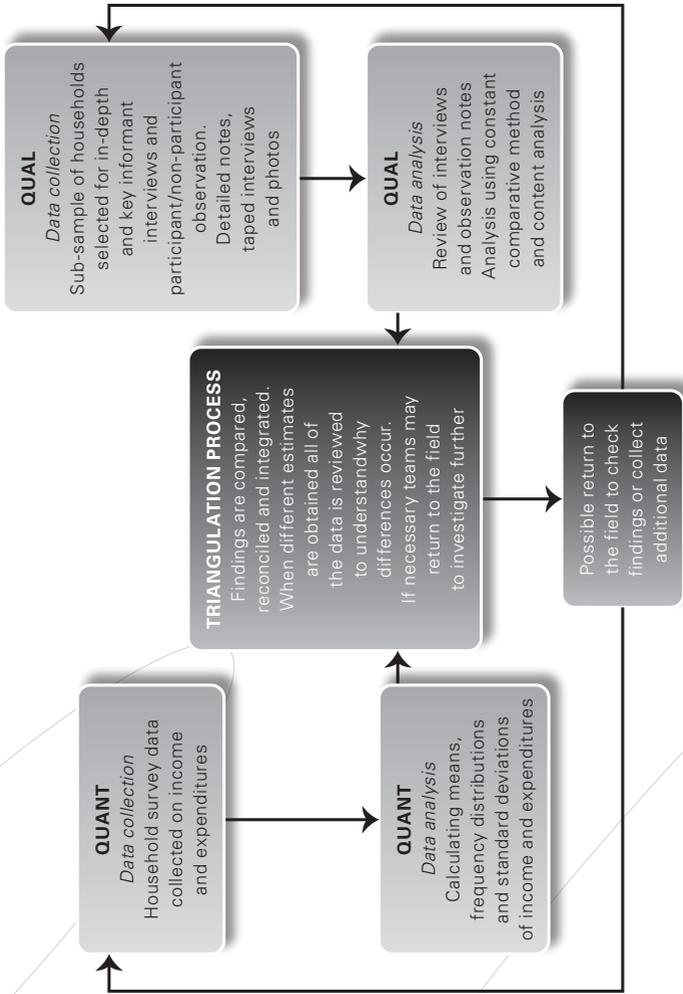
different interpretations of what actually happened during project implementation and what the effects were on different sectors of the population. Triangulation should be an integral component of the Equity-focused evaluation design and should be used to check the validity of the key indicators of processes, outputs and outcomes that are collected. Triangulation can involve:

- Comparing information collected by different interviewers.
- Comparing information collected at different times (of day, week, or year) or in different locations.
- Comparing information obtained using different data collection methods.

Figure 6 illustrates a strategy for building triangulation into the design of a study to assess changes in household economic conditions. QUANT data on the economic status of the household is collected through household surveys, and this data is analyzed using standard QUANT data analysis methods. At the same time QUAL data on the household economic status is collected from a sub-sample of households that have been selected from the main QUANT sample. QUAL data is collected using observation, in-depth interviews, focus groups and the analysis of household artifacts, and is analyzed qualitatively. The findings of the QUANT and QUAL analysis of household economic conditions are compared. If there are inconsistencies, the evaluation team meets to discuss possible explanations of the inconsistencies. If the reasons for the inconsistencies are not clear both teams will reanalyze their QUANT and QUAL data and meet again. If the reasons for the inconsistencies are still not clear, ideally one or both teams will return to the field to collect additional data so as to try to explain the inconsistencies.

Triangulation can also be used to obtain different perspectives on what actually happened during project implementation and what effects the project had on different groups. This can be done through interviews with individuals, focus groups, review of project documents or participant observation.

Figure 6: Validating findings through triangulation:
Estimating household income and expenditures



B. Attribution, contribution and the importance of the counterfactual

When evaluating the effects of development interventions it is important to distinguish between: *changes* that have taken place in the target population over the lifetime of the intervention, and *impacts* that can reasonably be attributed to the effect of the intervention. Statistical impact evaluations estimate the size of the change in the

project population (the *effect size*⁹), and the statistical probability that the change is due to the intervention and not to external factors. Many evaluations, particularly those conducted under budget and time constraints, only measure changes in the target population and *results* are often discussed as if they prove causality. It is important to appreciate that *change does not equal causality*. Interventions operate in a dynamic environment where many economic, social, political, demographic and environmental changes are taking place and where other agencies (government, donors, NGOs) are providing complementary or competing services, or introducing policies, that might affect the target population.

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*. In statistical evaluation designs (experimental and quasi-experimental), the counterfactual is estimated through a comparison group that matches the target population. If the comparison group is well matched, and if the level of change between this and the target group is sufficiently large to be statistically significant, then it is assumed that the difference is due, at least in part, to the effect of the intervention.

In the real-world, it has only proved possible to use statistical comparison groups in a small proportion of interventions, so evaluators have had to use their creativity to define *alternative counterfactuals*. This is particularly the case for policy interventions and other multi-component programmes, where it is rarely possible to use a statistical comparison group. In addition, a weakness of many statistical evaluation designs is that when expected outcomes are not achieved, it is difficult to know whether this is due to weaknesses in the underlying programme theory and how it is translated into project design (design failure), or whether it is due to problems with how the project was implemented (implementation failure). Economists often call this the “black box” problem, because pro-

9 The effect size is the average difference in the change score in the outcome indicator between the treatment and comparison groups. In pre-test/post-test comparison group design the change score is the difference in the mean pre-test/post-test scores for the project and comparison groups. In a post-test comparison group design the change score is the difference between the means of the two groups. If a single group design is being used in which the project group is compared with an estimate for the total population, the change score is the difference between the mean of the project group and that of the total population. Ideally the difference should be divided by the standard deviation of the change to estimate the size of the standardized effect.

ject implementation is a mysterious black box that is not analyzed and whose effects are not understood. The “black box” has many practical implications, because many clients assume that if an evaluation does not detect any statistically significant project impacts this means the project should be terminated, whereas often the recommendation should have been to repeat the project with more attention to how it is implemented.

Therefore, one of the main challenges for Equity-focused evaluations is how to define a credible counterfactual to answer the question “what would have been the situation of the worst-off groups if the intervention had not taken place”? Based on the above, one of the best ways to define credible counterfactual in Equity-focused evaluations is through contribution analysis.

Contribution analysis is used in contexts where two or more donor agencies, as well as one or more national partners, are collaborating on a programme or broad policy reform, and where it is not possible to directly assess the effects of a particular donor on overall outcomes and impacts. Sometimes contribution analysis for a particular donor will be complemented by attribution analysis, assessing the overall outcomes and impacts of the collaborative programmes (for example a poverty reduction strategy), but in most cases no estimates will be made of overall programme outcomes. The purpose of contribution analysis is to assess the contribution that a particular international agency has made to achieving the overall programme objectives.

The simplest form of contribution analysis is to define each stage of the programme (consultation; planning; design; implementation; achievement of outputs and outcomes; dissemination of findings; and sustainability) and to assess the agency’s contribution to each stage. The assessment combines a review of project reports and other documents¹⁰ with interviews with other international and national agencies and key informants. Interviews are often open or semi-structured but for large programmes rating scales may be used to assess performance on each component, as well as to assess the agency on dimensions such as collaboration, flexibility (for example with respect to use of funds), promoting broader participation etc. Agencies can also be rated on what they consider

10 Publications, planning documents, and meeting minutes of national planning agencies or line ministries can provide a useful source of information on how these agencies perceive the contribution of different partners. For example, if a donor believes it had a major influence on policy reform it can be interesting to see whether there are references to this in documents such as the Five Year Plan.

to be their areas of comparative advantage, such as knowledge of the national or local context, ability to work with a broader range of actors, or technical expertise.

John Mayne (2008)¹¹ proposes a theory-based approach to contribution analysis that includes the following steps:

1. Set-out cause and effect issues to be addressed in the analysis.
2. Develop the assumed theory of change and assess the risks to the achievement of the proposed changes.
3. Gather the existing evidence relevant to the theory of change.
4. Assemble and assess the contribution story (what changes took place, why did they take place and what were the contributions of the agency) as perceived by the agency being studied and by other partners. Identify and assess challenges to this story (for example some stakeholders or informants may not accept the claims made by the agency about their role in the changes).
5. Seek out additional information that both supports, and if necessary, challenges the contribution story.
6. Revise and strengthen the contribution story.
7. In complex settings, assemble and assess the complex contribution story.

When using the analysis to assess contributions to the achievement of equity objectives, each stage of the analysis must focus on equity-issues, using the kinds of questions discussed earlier.

C. Equity-focused evaluation at the policy level

The design of an equity-focused evaluation will depend on the nature of the interventions to be evaluated: national policy, programme or project.

While designing a project-level evaluation does not imply particular methodological challenges, it becomes more difficult to evaluate complicated equity-focused programmes using conventional evaluation designs. Sometimes conventional evaluation designs are applied to individual components of the programme and the overall programme performance is assessed by combining findings from

11 John Mayne 2008. *Contribution analysis: An approach to exploring cause and effect*. Rome: Institutional Learning and Change Initiative, ILAC Brief No. 16. May 2008. http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_contribution_analysis.pdf

the different components with other broader assessments of management, accessibility to the target population etc. When there is a systematic design for determining which individuals or organizations (schools, clinics etc.) receive which services, it may be possible to use a multivariate design that assesses overall outcomes and then assesses the contribution of each main component.

As described below, 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-level Equity-focused evaluations.

This section presents selected approaches to evaluate equity-focused interventions at policy level.

*Systems approaches to evaluation*¹²

Most development agencies, including UNICEF, are seeking to improve the welfare of the worst-off groups of society through finding the most effective way to deliver services to these groups, or to improve the performance of national policy, planning and service delivery agencies in reaching and benefiting these groups. All of the development interventions operate in, and often attempt to change, public and private service delivery systems and national governance and policy systems. All of these systems involve many actors and stakeholders, and often involve interventions with many stages. In addition, they operate through, and are affected by, other parts of the system. Interventions are also introduced into systems that have historical traditions (including perceptions about what will and will not work) and traditional ways of doing things. The interventions are also influenced by a wide range of economic, political, organizational, legal, socio-cultural and environmental factors. Finally, many programmes also involve the value systems of different actors concerning the target populations and what programmes and approaches should and should not be introduced.

Most conventional approaches to evaluation tend to address development programmes as largely stand-alone interventions, sometimes including contextual variables as factors affecting, but not really part of, the programme delivery system. Systems approaches have been developed to analyze these kinds of complexity and they

12 This section is based on Williams, B. (2005), *Systems and Systems Thinking* in Mathison, S. (editor) Sage Encyclopedia of Evaluation pp. 405-412. For examples of how these approaches are applied in practice see Williams, B. and Imam, I. (Eds.), (2007), *Systems Concepts in Evaluation: An expert anthology*. American Evaluation Association.

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 programme implementation takes place. Systems approaches can be particularly helpful for evaluating equity-focused policies as many of these operate within, and seek to change, systems which are often resistant to (or are actively opposed to) accepting the proposed focus on the worst-off groups in society.

Systems thinking introduces some radically different ways of thinking about evaluation, all of which are potentially important for Equity-focused evaluation. Some of the ideas, that can be drawn from approaches such as those described above, include:

- Programmes, policies and other kinds of development interventions are normally embedded in an existing social system that has its own historical traditions, linkages among different customers (clients/beneficiaries), actors and owners. The intervention must adapt to the existing system and will often be changed by the system.
- Different actors who may have very different perspectives on how the new intervention operates, and even whether it is accepted at all, will be affected in different ways by these perspectives.
- Systems have boundaries (which may be open or closed), which will often affect how widely the new intervention will be felt.
- New interventions create contradictions and often conflicts and the programme outcomes will be determined by how these conflicts are resolved.

Box 4. An example of a systems approach to Equity-focused evaluation: Evaluating the impact of social assistance on reducing child poverty and child social exclusion in Albania

This evaluation utilizes a simple form of systems analysis to assess how effectively the public sector policy and service delivery systems are able to reach the poorest and most vulnerable families and the implications this has for the access of children in these households to the social security provisions provided by the state.

For more information see: http://www.unicef.org/evaldatabase/index_59597.html

It is not possible to summarize the many different systems thinking methods in the present document but the following three approaches illustrate some of the approaches that could potentially

be applied to Equity-focused evaluations. While systems theory often has the image of being incredibly complex, in fact the goal of many approaches, including those described below, is to simplify complex systems down to their essential elements and processes. We illustrate how each of the three approaches could be applied to the evaluation of pro-equity service delivery systems.

The System Dynamics approach

This approach focuses on a particular problem or change that affects the ability of the system to achieve its objectives. It examines the effect of feedback and delays, how the system addresses the problem, and how the different variables in the system interact with each other, and how the effects vary over time. The focus is on system dynamics, adaptation and change rather than on a descriptive snapshot of the system at a particular point in time.

Applying this approach to evaluating the delivery of equity-focused services (e.g., adapting a current programme aiming to provide pre- and post-natal services to overcome resistance to extending services to vulnerable mothers and their children). The System Dynamics approach would study the way in which the new services were delivered; reactions of targeted mothers (feedback); how this affected the way the services were delivered; the effects of delays in implementing the new services, on the attitude and behavior of different actors, and on the effectiveness of reaching the target population. It would also examine how the introduction of the new service delivery mechanism affected the overall operation of the service delivery system.

Soft Systems Methodology

Soft Systems Methodology focuses on the multiple perspectives of a particular situation. The first step is to provide a “rich picture” of the situation and then to provide a “root definition” (the essential elements) of the situation in terms of:

- the beneficiaries;
- other actors;
- the transformation process (of inputs into outputs);
- the world-views of the main actors;
- the system owners (who have veto power over the system); and,
- environmental constraints.

Once the root definition has been defined a cultural analysis is conducted of the norms, values and politics relevant to the definition.

One or more system models are then defined using only the elements in the root definition (an example of how the systems approach seeks to simplify the system). A key element of the approach is that a new root definition can then be defined based on the perspectives and values of a different customer, actor or owner.

Applying this approach to evaluating the delivery of equity-focused services. A "rich picture" (detailed description) of the programme would be developed covering the 6 elements of the root definition. The service delivery system would be examined from the perspective of different elements of the worst-off groups, the different actors and owners. Areas of consensus as well as disagreement or conflict would be examined. Particular attention would be given to the attitude of the different "owners" who have the power to veto the new service delivery systems.

Cultural-Historical Activity Theory

The key elements of the Cultural-Historical Activity Theory approach are that:

- systems have a defined purpose;
- they are multi-voiced (different actors have different perspectives);
- systems are historical and draw strongly from the past;
- changes in a system are produced largely by contradictions which generate tensions and often conflict; and,
- contradictions provide the primary means by which actors learn and changes take place. The changes can produce further contradictions so processes of change are often cyclical.

Applying this approach to evaluating the delivery of equity-focused services. Actors have different perspectives on whether and how services should be extended to vulnerable groups. These different perspectives – combined with the fact that the changes required to address the needs of worst-off groups can create contradictions – and how these are resolved, will determine how effectively the services reach worst-off groups. The Cultural-Historical Activity Theory approach also stresses that the cyclical nature of processes means that the changed procedures will often result in a cyclical process with further revisions, so that short term success in reaching vulnerable groups should not be assumed to be permanent.

Unpacking complex policies

Many complex policies and other national-level interventions have a number of different components each with different objectives and organized in different ways. Many agencies conclude that most of these interventions are too complicated for a rigorous evaluation to be conducted, or to use any of the conventional comparison group designs. Also, as the interventions are defined at the national level and are intended to operate throughout the country, it is assumed that it is not possible to find a comparison group that is not affected. However, it is often possible to “unpack” the policy into a number of distinct components, making it possible to design a more rigorous evaluation:

- Complex policies can often be broken down into different components, each with clearly defined structures and objectives.
- While policies are formulated at the national level, in many cases they will be implemented and will have measurable outcomes at provincial and local levels.
- Even though policies are intended to cover the whole country, they tend to be implemented in phases, or for different reasons 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 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 (but that are scheduled to be covered by future phases) can be used as a comparison group. While there are many situations in which policies are implemented in phases and where pipeline designs can be used, 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 pipeline design may work well, but when certain regions have been unintentionally excluded due to problems (administrative or political) the use of the pipeline design may be more problematic¹³.

13 Very often the excluded regions or areas are poorer, or government agencies have more limited administrative capacity (often due to more limited resources), so there will often be systematic differences between them and the areas where the policies are being implemented – limiting their validity as a comparison area.

Policy gap analysis

Policy gap analysis is a term used to describe analytical approaches that identify key policy priorities and target groups and assess how adequately current and planned policies address these priorities. It reviews the whole spectrum of public sector policies to identify both limitations of individual policies and also 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. In Central and Eastern Europe, for example, UNICEF adopts a systemic approach to the assessment of the adequacy with which countries address issues of vulnerability as they affect children and their families¹⁴.

The analysis is normally conducted at the national level although it can also be applied in a particular region or sector. The analysis normally relies on secondary data from surveys and agency records. Techniques such as quintile analysis are used to identify the worst-off groups and to compare them with other groups through indicators such as school enrolment or use of health services¹⁵. If available, studies such as Citizen Report Cards can provide additional useful information.

Often these secondary data sets do not include all of the required data (for example they may not cover both supply and demand-side factors), in which case they may be complemented by other data sources such as records from public service agencies. Techniques such as Bottleneck Analysis or Knowledge, Attitude and Practice (KAP) studies could make a major contribution to the data requirements for policy gap analysis. It is sometimes possible to develop a special module that can be incorporated into an ongoing or planned survey to fill in some of the information gaps. These data sources will normally be complemented by desk reviews, consultation with key informants, focus groups and possibly visits to ministries or service delivery centres.

14 See Albania: Evaluating the impact of social assistance on reducing child poverty and social exclusion

15 When the quality of secondary data permits, it is possible to use techniques such as social exclusion analysis and multidimensional poverty analysis to identify vulnerability in terms of a much wider range of indicators

Using other countries or sectors as the comparison group

For policies that are implemented country-wide or that cover all of the activities of a ministry, one option is to use other countries as a comparator. One or more countries can be selected in the same region. In these cases, it is difficult to use a statistical comparison and the analysis will normally be descriptive, drawing on whatever kinds of comparative data are available. As each country is unique a great deal of interpretation and judgment will be required.

A second option is to draw on the increasingly rich international databases now available. Extensive comparative data is available for most of the MDGs, household socio-economic and demographic conditions, human development indicators, and access to public services. Over the past few years databases are also becoming available on governance and participatory development topics such as corruption and political and community participation.¹⁶ These databases permit the selection of a large sample of countries with similar socio-economic and other relevant characteristics. Changes in key outcome indicators for the target countries are then compared with other similar countries that have and have not introduced reforms. It is however more difficult to find data relating to worst-off groups and where data is available it will normally apply to income comparisons and will not address other dimensions of inequity.

Sometimes, when a policy is being launched in different ministries or agencies, it may be possible to use as the comparison the ministries where the programme has not yet started. Policy areas where this type of comparison could be considered include: anti-corruption and other kinds of administrative reform, decentralization and financial management. However, these comparisons are difficult to apply as every agency has unique characteristics. Also, it is difficult to obtain baseline data on the situation before the reforms began as information on outcome indicators tends to be limited, not very reliable and difficult to compare with the extensive and more rigorous indicators that the reform programmes tend to generate.

16 Pickett and Wilson (2009), *The spirit level*; Hills, Le Grand and Piachaud (2001), *Understanding social exclusion*; and the UNDP Human Development Index provide examples of the wide range of secondary data sources that are now available for assessing the causes and consequences of vulnerability.

Concept mapping

Concept mapping uses interviews with stakeholders or experts to obtain an approximate estimate of policy effectiveness, outcomes or impacts. It is well suited as a tool for Equity-focused evaluation as it allows experts to use their experience and judgment to help define the equity dimensions that should be used to evaluate policies, and then to rate policies on these dimensions. This is particularly useful for the many kinds of 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. The approach is described in “Using concept mapping to evaluate equity-focused policy interventions” available in www.mymande.org and an illustration is given of how this could be used to assess the effectiveness and impacts of a gender mainstreaming strategy being implemented in different countries. A similar approach could be applied to evaluate a wide range of equity-focused policies that seek to increase access by 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, particularly when supported by several different stakeholders, can include large numbers of different interventions. An equity-focused example would be a gender mainstreaming and women’s empowerment programme that might include components from large numbers of different policy and programme interventions, including many where gender mainstreaming was only one of several objectives. Portfolio analysis is an approach that is commonly used in these cases. All interventions are identified (which can in itself be a challenge) and then classified into performance areas. A desk review is then conducted to check the kind of information that is available on these projects such as: the existence of a logic model; monitoring data on inputs and outputs; ratings of quality at entry; quality of implementation; quality at completion; and other kinds of evaluation reports. Often there is no clear delineation of the projects to be included in the analysis, and boundary analysis may be required to define criteria for determining which projects should and should not be included.

If the information on each project is sufficiently complete, which often is not the case, projects will be rated on each dimension and

summary indicators will be produced for all of the projects in each performance area. For example, quality at entry or during implementation may be assessed in terms of: quality of design; quality of planning; the design and use of the M&E system; and the internal and external efficiency. Where data permits, average ratings will be computed for each of these dimensions and an overall assessment will be produced for quality of entry or implementation. The ratings for the different components (quality at entry etc.) are then combined to obtain an overall assessment for each performance area. Many agencies use the OECD/DAC evaluation criteria for these overall assessments. Additional criteria relevant to humanitarian settings, such as coherence, connectedness and coverage, may also be used.

If resources permit, a sample of projects from each performance area will be selected for carrying-out field studies to compare the data from these secondary sources with experience on the ground. The findings will then be reviewed by a group of experts and stakeholders, and where there are discrepancies between the draft reports and the feedback from this group, further analysis will be conducted to reconcile or explain the reasons for the discrepancies. In some cases the kinds of concept mapping techniques described earlier in this chapter may be used as part of the assessment. In cases where field studies are conducted, concept mapping can also be used to help select the countries or projects to be covered.

D. Equity-focused evaluation at the project and programme levels

Conventional quantitative impact evaluation designs

Project-level impact evaluation designs estimate the contribution of an intervention (project) to the observed changes in an outcome indicator (the change the project seeks to produce). This is done by identifying a comparison group with similar characteristics to the project population, but that has no access to the intervention. The comparison group serves as the control for changes due to external factors unrelated to the project. Table 3 represents a pre-test/post-test comparison group design. P_1 and P_2 represent the measurements (surveys, aptitude tests, etc.) taken on the project (treatment) group before and after the project (treatment) has been implemented. C_1 and C_2 represent the same measurements on the comparison group at the same two points in time. If there is a statistically significant difference in the change that occurs in the project group, compared to the change in the comparison group, and

if the two groups are well matched, then this is taken as evidence of a potential project effect. The strength of the statistical analysis is influenced by how closely the project and comparison groups are matched, as well as the size of the sample and the size of the change being estimated (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. Ideally the impact evaluation should be repeated several times on similar projects (as in laboratory research) but this is rarely possible in the real world.

Table 3: Conventional pre-test/post-test comparison group impact evaluation design

	Pre-test	Project intervention	Post-test
Project (treatment) group	P ₁	X	P ₂
Comparison (control) group	C ₁		C ₂

Notes: 1. The comparison group is used to define the counterfactual “what would have been the condition of the project population if the project had not taken place?” The strength of the counterfactual depends on how well the project and comparison groups were matched.

2. A major determinant of the statistical strength of the evaluation design depends on how the comparison group was selected (see following text).

The statistical validity of the estimate of project effect (impact) is affected by how well the project and comparison groups are matched. The three main methods for matching, in descending order of statistical precision are:

- *Randomized control trials* in which subjects are randomly assigned to the project and control groups.
- *Quasi-experimental designs*¹⁷ in which secondary data permits the comparison group to be *statistically matched* with the project group.

17 A quasi-experimental design (QED) is an evaluation design in which project beneficiaries are either (a) self-selected (only people who know about the project and chose to apply/participate) or (b) participants are selected by the project agency or a local government agency in the location where the project will be implemented. In either case the project beneficiaries are not an unbiased sample of the total target population and in most cases the people who enter are likely to have a higher probability of success than the typical target population member. The evaluator then tries to select a comparison group sample that matches as closely as possible the project beneficiaries.

- *Quasi-experimental designs* in which *judgmental matching* is used to select the comparison group.

There are a large number of design options that can be considered, but the range of viable options is often limited by budget, availability of secondary data, and when the evaluation began. In the Equity-focused evaluation resource centre a list of 7 basic impact evaluation designs is presented based on: when the evaluation began (start, middle or end of the project); whether there was a comparison group, and whether baseline data was collected on the project and/or the comparison group.

An expanded list with 20 evaluation design options is also presented. This builds on the 7 basic designs but also takes into consideration two sets of factors. Firstly, whether the comparison group (counterfactual) was selected randomly, using a quasi-experimental design with statistical matching or judgmental matching, or whether the counterfactual was based on a qualitative design. Secondly, how was the baseline condition of the project and comparison groups estimated: conducting a baseline survey at the start of the project; “reconstructing” the baseline condition when the evaluation is not commissioned until late in the project cycle; using qualitative methods to estimate baseline conditions; or, no information is collected on the baseline condition of the project and comparison groups.

Estimating project impacts using non-experimental designs

Non-experimental designs do not include a matched comparison group (statistical counterfactual) so it is not possible to control statistically for the influence of other factors that might have produced the changes in the output indicators. It is useful to distinguish between situations where a non-experimental design is used as the default option, because time and resource constraints do not permit the use of a comparison group; and situations where, in the judgment of the evaluators a non-experimental design is the methodologically strongest evaluation that can be used. Situations where non-experimental design might be considered the best design include:

- When the project involves complex processes of behavioral change that are difficult to quantify.
- When the outcomes are not known in advance, as they will either depend on the decisions of project participants or on interactions with the other actors.

- When many of the outcomes are qualitative and difficult to measure.
- When each project operates in a different local setting and where elements of this setting are likely to affect outcomes.
- Where there is more interest in understanding the implementation process than in measuring outcomes.
- Where the project is expected to evolve slowly over a relatively long period of time.

Potentially strong non-experimental designs

Some of the potentially strong non-experimental designs that could be considered include:

- *Single case analysis.* This is a pre-test/post-test comparison of a single case (such as a child suffering from behavioral problems in a classroom). The baseline observation, before the treatment, is taken as the counterfactual. The treatment is applied at least three times, and if a significant change is observed on each occasion (usually based on the observation ratings of experts) then the treatment is considered to have been effective. The experiment would then be conducted again in a slightly different setting to gradually build up data on when and why it works.
- *Longitudinal designs.* The subject group, community or organization is observed continuously, or periodically over a long period of time, to describe the process of change and how this is affected by the contextual factors in the local setting. One option is to select a small sample of individuals, households or communities who are visited constantly over a long period of time (panel study). This approach is useful for understanding behavioral change, for example in relations between spouses as a result of a programme to promote women's economic empowerment. It has been used successfully to evaluate, for example, the effects of microcredit programmes on women's empowerment. A second option is to observe the group or community over a long period of time, to monitor, for example, changes in the level of gender-based violence in the community or market.
- *Interrupted time series.* The design can be used when a series of observations at regular intervals is available over a long period of time, starting well 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. This method has been widely used to evaluate, for example, the impact on the number of road accidents of new anti-drinking legislation.

- *Case study designs.* A sample of case studies is selected to represent the different categories or typologies of interest to the evaluation. The typologies may be defined on the basis of quantitative analysis of survey data or they may be defined from the qualitative diagnostic study. The cases describe how different groups respond to the project intervention and this provides an estimate of project impacts.

E. Feasibility analysis

Once the evaluation design has been proposed it is important to assess its feasibility. This involves questions such as: Can the data be collected? Will it be collected within the budget and time constraints? Can the design address the key evaluation questions? Will the evidence be considered credible by key stakeholders? The feasibility analysis must also assess the credibility of the proposed counterfactual – particularly when non-experimental designs are used.

An important issue, that weakens the validity of the findings of many evaluation designs, is the point in the project cycle at which the evaluation is conducted. Due to administrative requirements and pressure to show that the project is achieving its intended outcomes, many impact evaluations are commissioned when it is still too early in the project cycle to assess outcomes. For example, it may require several years before a girls' secondary education project can have an effect on age at marriage or teenage pregnancies; but due to donor or government pressure the evaluation may be conducted at the end of the first year before the programme has had an effect.

5.3 Collecting and analyzing data

The evaluation manager must ensure that fieldwork meets evaluation method standards for gathering evidence to support findings and recommendations on the intervention's contribution to equity. Defining the tools for data collection and analysis is the first part of implementing a successful evaluation process. The next section describes some of the tools appropriate for Equity-focused evaluations. In addition to being robust and generating reliable data,

the tools selected should maximize the participation of stakeholders identified in the stakeholder analysis, allowing for active, free, meaningful participation by all.

A. Collecting data and analyzing contextual factors

When designing an Equity-focused evaluation it is important to understand the context within which the intervention has been implemented, and the factors that affected implementation and accessibility to the different worst-off groups. It is also important to understand the perceptions and attitudes of implementing agencies and society towards the different worst-off groups.

In most situations it will be useful to conduct a rapid diagnostic study to understand the intervention and its context. The type of study will be determined by the size and complexity of the intervention; how familiar UNICEF and its partners are with this type of intervention and with the locations where it will be implemented. For a small intervention implemented 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 usually cover:

- How are problems the intervention is designed to address, currently being addressed? Do other agencies provide these services? Are there traditional approaches for addressing the problems?
- What are the opinions of different sectors of the community concerning these services? Who uses them and who does not?
- Have similar projects been tried earlier? How did they work out? Why were they discontinued?
- Which groups are most affected by the problems to be addressed? Would they be considered as worst off, and if so in which category would they be classified?
- What are the reasons for lack of access of different groups to the services? How would these be categorized in the bottleneck framework?
- Are there any cultural attitudes or practices that affect access to the planned services and how they are used – particularly by worst-off groups?

Box 5. Analysis of contextual factors affecting the outcomes of equity focused evaluations

The following evaluation case studies (see section 8) illustrate different ways in which contextual factors affected the implementation and outcomes of equity-focused programmes:

- Evaluating the UNICEF education programme in Timor l'Este. How conflict, extreme poverty, and the breakdown of government affected access of vulnerable groups to schools.
- Evaluating the Education for All Programme in Nepal. The effects of geographical remoteness and ethnicity on access to education.
- Evaluation of the international humanitarian response to the 2009 displacement crisis in Pakistan. The effects of military and government control on access to conflict areas on the delivery of emergency services to the displaced population.
- Evaluation of the Community Justice Facilitation Project in Tanzania . The effects of locations, the limited resources available to local government and the limited attention to gender in the access of vulnerable populations (particularly girls) to community justice.

Diagnostic studies will normally use one or more of the following data collection methods:

- *Participant observation*¹⁸. One or more researchers live in the community or become involved in the group or organization as participating members or as people who are known and trusted. The goal is to live the experience of the project and of living in the community in the same way as other residents, rather than simply observing as an outsider. It is important to be aware of the ethical implications in cases where the researcher does not fully explain who s/he is and why s/he is living in the community or participating in the group.
- *Non-participant observation*. Many kinds of observation are possible without having to become accepted as a member of the community. For example: it is possible to observe how

18 For a detailed description of the participant observation approach see Salmen, L (1987) *Listen to the People: Evaluation of Development Projects*. Salmen lived for 6 months in low-income urban communities in Bolivia and Ecuador to experience the first World Bank low-cost urban housing programmes in the same way as community residents. By living in the community and winning the trust of residents he was able to discover many critical facts that previous evaluation studies had failed to capture. For example, he found that there was a very large undocumented renter population who became worse off as a result of the project, but neither previous researchers nor project management were aware of their existence as they hid whenever outsiders came to the community as they were afraid they would be evicted.

water and fuel are collected, transported and used and the kinds of conflicts and problems that this causes; and the use and maintenance of social infrastructure such as community centres, schools, drainage channels, and children's playgrounds. A lot can be learned by watching people entering health centres, village banks and schools¹⁹. It is important to recognize that the presence of an outsider in the community will change behavior however inconspicuous they try to be. For example, drug dealers may move elsewhere and residents may not engage in informal business activities not permitted by the housing authority.

- *Rapid household surveys.* If the number of questions is kept short, it is often possible to conduct a large number of interviews in a relatively short period of time. For collecting information on hard-to-reach groups or excluded groups, it is generally better to use people from the community or from local organizations. It is of course necessary to ensure that local interviewers have the necessary experience and credibility.
- *Key informants.* Key informants are a valuable source of information on all of the questions mentioned above and for understanding relations within the community and between the community and outside agencies (government, private sector and NGOs). Key informants are not only government officials, academics, religious leaders and donor agencies, but also representatives from worst-off groups and in general, anyone who has extensive knowledge on the questions being studied. Teenagers will be a principal source of information on why teenagers do, and do not, attend school. Key informants always present information from a particular perspective, so it is important to select a sample of informants who are likely to have different points of view to counterbalance each other. The use of triangulation is important when attempting to reconcile information obtained from different informants.
- *Local experts.* These are people who are likely to have more extensive and credible knowledge on topics such as health statistics, availability of public services, crime, school attendance and overall economic conditions. However, many experts may have their own perspectives and biases which must be taken into consideration. For example, the local police chief may be

19 In one assessment of women's access to rural health centres it was observed that women using traditional dress seemed to be treated less well than women in western dress.

anxious to prove that crime rates have gone down since s/he was appointed (or that crime has increased if s/he is seeking support for a budget increase!).

- *Focus groups.* Groups of 5-8 people are selected to cover all the main groups of interest to a particular study. For example, each group might represent a particular kind of farmer or small business owner; poorer and better off women with children in primary school; men and women of different ages, and perhaps economic levels, who use public transport. While most focus groups select participants who come from the same category of interest group to the study, another strategy is to combine different kinds of people in the same group²⁰.

Box 6. Focus groups are not a fast and cheap way to collect information.

Focus groups have been widely abused by evaluators who look on this as a fast and cheap substitute for a survey. A well designed focus group requires a lot of preparation and careful selection of a representative sample of participants, as well as considerable time for analysis and reporting findings. Calling a local government agency or an NGO the day before to “invite a group of mothers to talk about child health-care issues” is not a focus group but only an informal and usually unstructured conversation.

A useful reference source is Krueger, R. and Casey, M. (2000), *Focus Groups: A Practical Guide for Applied Research*.

B. Collecting and analyzing information to understand knowledge, attitude and practices

Knowledge, attitude and practices information on public services should be collected in relation to different groups:

- *Worst-off groups.* Information is needed on their understanding of the nature of health and other problems and the actions they must take to address these problems. Worst-off groups suffer from multiple problems so that taking actions, such as coming to a clinic or detox centre, or acquiring and using contraceptives, can be difficult and in some cases dangerous.
- *Service delivery agencies.* Information needs relate to their attitudes and how they interact with worst-off groups. There are a

20 In a study of reasons why female college students did not use public transport in Lima, Peru; some focus groups were conducted with homogenous groups, such as college-age girls, teenage boys, mothers etc., while others mixed teenage boys and girls and adult men and women. The attitudes to condoning sexual harassment on buses was very different in single sex and mixed groups.

wide range of knowledge gaps and ingrained attitudes (including fear and distrust, and feelings of superiority) that affect their adoption of open and supportive behavior.

- *Policy-makers and planners* who often make assumptions about worst-off groups, the causes of their problems and how they will respond to the provision of services. Often attitudes are based on “factoids” which are assumptions and bits of knowledge widely believed to be true, but which are often false or only partially true.

The information on attitudes and beliefs, and the evaluation of the effectiveness of different interventions in changing them, can be collected through *Knowledge, Attitude and Practice (KAP) studies*, using the following questions:

- *Knowledge*: was information about the intervention disseminated? Did it reach all groups of the target population, including worst-off groups, and was it understood?
- *Attitudes*: what did people, including worst-off groups, think about the new programmes or information? Did they agree with it or not?
- *Behavior (Practice)*: did they change their behavior? If they agreed with the information/programme did they adopt it? Was it properly implemented? If they did not adopt it, why was this: was it due to lack of access, to the attitudes or behavior of other household members, or to contextual factors?

Figure 8 illustrates the framework of a KAP study assessing the effectiveness of a campaign to introduce bed-nets to reduce malaria among pregnant women. The questions about practice are similar to the questions in the bottleneck analysis about demand for services and effective utilization.

The five steps in designing a KAP study are the following:

Step 1: *Domain identification*: defining the intervention, the knowledge to be communicated, the attitudes to be measured and the indicators of acceptance and use.

Step 2: *Identifying the target audience*: in the example of bed-nets it would be necessary to decide whether the campaign is just targeted at pregnant women, or also at other family members, or other members of the community (such as traditional birth attendants), and perhaps local health professionals.

Step 3: *Defining the sampling methods:*

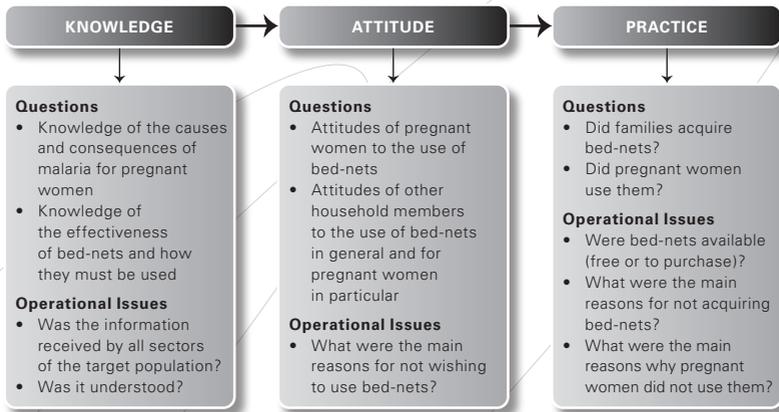
- Defining the population to be sampled: the geographical areas and the target populations.
- Defining the sample selection procedures: often the population will be broken down into sub-groups, each with special interests or issues. If worst-off groups who are difficult to identify and to reach are targeted, special sampling procedures might be required, such as, snowball sampling; quota sampling; multi-stage sampling; requesting assistance from key informants or group leaders; sociometric techniques; and, identifying people in locations known to be frequented by the targeted worst-off groups.

Step 4: *Defining the data collection procedures:* ideally KAP studies should use a mixed-method data collection strategy combining the following types of quantitative and qualitative data collection methods:

- Sample surveys.
- Observation (participant or non-participant).
- Key informant interviews.
- Focus groups.
- Inclusion of questions in an omnibus survey questionnaire already planned.
- Project records.
- Secondary data sources such as reports and records from other agencies and previously conducted surveys.

Step 5: *Analysis and reporting:* this follows standard practices for survey analysis and reporting on focus groups, key informants and observation studies.

Figure 7. Hypothetical example: KAP analysis of a campaign to introduce bed-nets to reduce malaria among pregnant women



C. Collecting and analyzing information on the quality of services delivered and the satisfaction of citizens

Citizen report cards²¹

Citizen Report Cards are based on large surveys that typically cover a major urban area (the first study was conducted in Bangalore, India). The survey asks households which public service agencies (education, health, police, transport, water etc.) they have had to contact within the last 12 months to address a particular problem. For each agency they are asked: were they able to resolve their problem; how many visits were required; how were they treated by agency staff; did they have to pay bribes (if so how many and how much). 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 Bangalore study included a separate stratum for slum dwellers). Studies can

21 For an example of a citizen report card study see Bamberger, MacKay and Ooi (2005), *Influential Evaluations: detailed case studies. Case study No. 3 Using Citizen Report Cards to Hold the State to Account in Bangalore, India*. Operations Evaluation Department. The World Bank. Available at: www.worldbank.org/oed/ecd

either cover all of the main public service agencies or they can just focus on a particular sector such as health or education.

Experience shows that the credibility and independence of the research agency is critical as a typical reaction of agencies is to claim that the findings are not representative and to challenge the professional competence or motives of the research agency. For the same reason it is important to have a sufficiently large sample to be able to disaggregate the data by different worst-off groups.

D. Carrying out cost-effectiveness studies to compare costs and results of alternative interventions

Cost-effectiveness analysis²²

Cost effectiveness is the defining element of a method for comparing both the costs and the results of different options for addressing particular goals. Criteria for measuring effectiveness must be similar among the different options for a cost-effective comparison. Effectiveness estimates are based on the usual experimental, quasi-experimental or statistical designs. Cost estimates are based on a careful specification of required resources and their market values. Selection of the options having the greatest effectiveness per unit of cost will generally provide the largest overall impact for a given resource constraint. In performing a cost-effectiveness analysis, adequate scrutiny must be given to both the cost measurement and the estimation of outcomes (Levin, 2005).

Cost-effectiveness analysis is used for comparing different services or delivery systems. It may involve a comparison between average costs of service delivery and costs for reaching special groups (e.g. worst-off groups), or it may involve comparisons between different 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. For example, it is not possible to compare directly the cost-estimates for a malaria control programme run by an NGO and involving orientation sessions and follow-up home visits, in addition to

22 Much of the work on cost-effectiveness has been conducted in the areas of education and health. For a good overview of cost-effectiveness methods see Levin, H and McEwan, P. (2011), *Cost-Effectiveness Analysis: Methods and Applications*. Second Edition. Most of the examples are drawn from education but it provides a good introduction to the general principles. For an introduction to the application of cost-effectiveness in health see Muennig, P. (2008), *Cost-effectiveness analysis in health: A practical approach*. Wiley Publications

the malaria treatment, with a government programme that only involves the handing out of bed-nets and tablets with no orientation or follow-up.

- Identifying all of the costs of the programmes being compared and ensuring that they are measured in an equivalent way, and that any hidden subsidies are identified and monetized. For example, some NGOs may obtain free services from volunteer doctors whereas the government programme includes the full cost of doctors. On the other hand an NGO may have to pay rent for their clinic whereas the government programme may be provided with the space in the local health centre at no charge.
- Ensuring that standard definitions are used to record the number of users. This is critical because the average (unit) cost is calculated by dividing the total cost by the number of people treated. So it is important to clarify, for example, whether a mother who brings her child for a check-up and is given malaria treatment by the nurse (even though she had not come for this purpose), is counted as a person who was treated for malaria prevention. In multi-service clinics, how this is defined can have a major effect on the average cost estimates.
- A final issue concerns the question of scaling-up. Many programmes start on a small scale and if they are considered successful it will then often be recommended that they should be replicated on a larger scale. However, it is difficult to estimate how scale-up will affect costs. While there may be economies of scale from working with a larger number of patients/clients, on the other hand the larger organizational effort will require additional administrative staff, and perhaps more expensive computer systems. So care must be taken when assuming that because a small programme is relatively inexpensive that the same will be true if the programme is replicated on a larger scale.

Public expenditure tracking studies²³

Public expenditure tracking studies (PETS) track the percentage of budget funds approved for front-line service delivery agencies such as schools and health clinics, that actually reach these agencies. The studies are important because in some cases it has been found

23 For an example of a PETS study applied to education in Uganda see Bamberger and Ooi eds., (2005), *Influential Evaluations: Detailed case studies. Case 7 Improving the delivery of primary education services in Uganda through public expenditure tracking surveys*. Independent Evaluation Group, World Bank. Available at www.worldbank.org/oed/ecd

that less than 20% of approved budget funds actually reach the schools or clinics. The studies have proved to be an effective advocacy tool, mobilizing the media, public opinion and intended beneficiaries, to pressure government to improve the delivery of funds. If data is available it would be possible to track the proportion of funds that reach the programmes targeted at worst-off groups.

The studies involve a very 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.

Public expenditure Benefit Incidence Analysis²⁴

Benefit Incidence Analysis (BIA) estimates the effectiveness with which public expenditure in sectors such as health and education reach worst-off groups. Normally the analysis focuses on access to services by income quintile as data is more readily available on these groups, and the analysis is rarely able to examine other dimensions of inequity (such as female-headed households, and families with physically or mentally disabled children). The analysis requires three types of data:

- Government spending on a service (net of any cost recovery fees, out of pocket expenses by users of the service or user fees);
- Public utilization of the service; and
- The socioeconomic characteristics of the population using the service.

The analysis can either be used at one point in time or it 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 strategy programmes (PRSP) but it could have other applications and is a potentially useful tool for Equity-focused evaluations. Ideally BIA should be considered as one of several tools used for Equity-focused evaluations, with the weaknesses in data on aspects such as quality, and utilization by different household members etc., being complemented with techniques such as bottleneck analysis or KAP studies.

24 For an introduction to BIA see Davoodi, Tiongson and Asawanuchit (2003), *How useful are benefit incidence analyses of public education and health spending?*

BIA assesses the proportion of the health or education expenditure that benefit particular groups of users, such as households in each income quintile. Two limitations of BIA for Equity-focused evaluation and particularly for focusing on children are: data is normally not available on the quality of services, and data is normally only available at the level of the household so that it is not possible to examine access by different household members. This is critical for Equity-focused evaluation as there will often be differences in the frequency with which boy and girl children are taken to the health clinic, or the frequency with which women and men use services.

SECTION 6: UTILIZING THE EVALUATION

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

6.1 Preparing the evaluation report and alternative forms of reporting

After the data collection process, evaluators will analyze the data and prepare the evaluation report.

It is good practice to discuss evaluation findings with stakeholders, including worst-off groups, before the preparation of the report. It is an opportunity to explain how their contributions were used, and to provide them with the chance to correct any inaccuracies and to clarify any doubts. This can be done in the form of a final workshop, and the selection of participants should refer back to the stakeholder analysis, including special attention to the worst-off groups, who can often be left out of discussions due to multiple kinds of constraints. To adequately ensure equity, the workshop needs to follow the lines which were, ideally, already adopted in the evaluation process: being as inclusive as possible, and creating adequate space for reflection and active, free and meaningful participation.

A good evaluation report will need to make sure that the information provided by participants during the evaluation process, including the final workshop, is duly captured with balanced perspectives and fair representation of different points of view. Findings and recommendations need to be formulated in detail, identifying to whom the recommendations are addressed and proposing concrete actions. The evaluation report is the most important resource for enabling the evaluator to reassert the importance of adequately addressing equity. Table 4 presents some guidance on how to formulate an evaluation report to adequately addresses equity.

A traditional evaluation report may not be sufficient to inform all the audiences of an evaluation. At this stage in the process, the evaluation team will have been informed about the different audiences, and their particular needs, by the Steering Committee and stakeholder analysis. For example, there may be illiterate groups, or stakeholders who do not speak the official language of the evaluation. Understanding these differences and needs is key to the

inclusion of these stakeholders in the process of understanding the evaluation findings, learning from them and supporting the implementation of the recommendations. The evaluation team/manager can devise forms of evaluation reporting that make use of alternative ways of depicting information through, for example, imagery, theatre, poetry, music, etc.

Table 4: Preparing the evaluation report*

Additional elements relevant to an Equity-focused evaluation report
Coverage of equity information. The report should correspond with the requirements in the ToR for information and findings on equity.
Stakeholder participation. The report should acknowledge how inclusive stakeholder participation, including worst-off groups, was ensured during the evaluation process.
Recommendations on equity. Are the conclusions adequately supported by the findings? Do the conclusions warrant recommendations, and are they appropriately targeted and specific, and likely to lead to appropriate action? If not, can they be made more relevant? Will it be possible to follow up on the recommendation to see if it has been implemented?
Limitations. Challenges to obtaining equity information or to addressing the issues appropriately should be included. Indicate the implications of not having data available, if this is the case. If data had been available, what would have been different in the evaluation? What would have been the gains in the process?
Lessons. Include lessons on equity, both related to the intervention itself, and also on how to integrate this dimension into the evaluation.

**This table is adapted from UNEG 2011*

6.2 Disseminating the evaluation and preparing a management response

Once the evaluation has been completed, the evaluation manager is bound by his/her organization's policies on dissemination. However, they should promote the fullest possible use of the equity dimensions of the evaluation among key stakeholders, including worst-off groups, within the UN systems and among colleagues. Methods and elements of a good dissemination plan include:

- *Providing barrier-free access to the evaluation products.* Is the language and format of the report accessible to all potential users, including worst-off groups? Is it easy to find and disseminate?

- *Ensuring the direct users identified in the planning phase use the evaluation for the original intended use.* Refer back to the initial discussions in the Steering Committee and the stakeholder analysis in order to assess to whom the evaluation should be disseminated. How should they be engaged and how can they contribute to dissemination? How can direct users take advantage of their own channels to disseminate the evaluation?
- *Identifying indirect users of the evaluation.* There may be other groups who would be interested in the findings and conclusions of this evaluation, such as evaluation networks; gender focal points; human rights bodies; and civil society organizations, which can use the lessons and data identified. This may mean national, regional, or global users. Can the evaluation manager and the members of the Steering Committee use their networks to inform these groups about the evaluation, or publicize the evaluation on an organizational website, or agree to links on other websites?
- *Developing good practices and lessons learned.* Since the systematic inclusion of equity in UNICEF evaluations is a recent emphasis, especially for work that is not specifically targeting equity, it could be useful to compare experience in this area with evaluation colleagues in the UN system.

The United Nations Evaluation Group (UNEG) Norms and Standards and the UNICEF Evaluation Policy recommend preparing a management response to all evaluations. A management response addresses recommendations, identifying who is responsible for their implementation and what are the action points and deadlines. Management responses are a practical means to enhance the use of the evaluation findings and conclusions to improve action. They “force” evaluators to be clear and straightforward in their recommendations. In the spirit of participation, stakeholders, including worst-off groups, should also participate in the decisions on how to respond to the evaluation, and agree on clear roles and responsibilities. All agreed responses should take into consideration the possible effects on equity.

SECTION 7: CONDUCTING EQUITY-FOCUSED EVALUATIONS UNDER REAL-WORLD CONSTRAINTS

Additional material on this section is available at the Equity-focused evaluations resource centre available at www.mymande.org

There are many textbooks and guidelines that provide useful pointers on how to conduct evaluations when there is an adequate budget, sufficient time, and a reasonable expectation that it will be possible to collect the required data (either from existing secondary sources or from new data collection). However, there is much less guidance available on how to conduct credible and methodologically sound evaluations when conducting evaluations under budget and time constraints and with difficulties of collecting the required data. This section addresses some of these challenges and the implications for Equity-focused evaluations.

7.1 Understanding the evaluation scenario

When planning and designing an evaluation it is important to understand the context in which it will be implemented. This involves understanding the following aspects:

- The purpose for which the evaluation was commissioned, the specific questions of interest to intended users, how the results will be used, and decisions to which the findings will contribute. It is also important to understand the critical deadlines for receiving the information. Given the sensitive nature of many equity issues, it is important to understand the expectations of different stakeholders and how they plan to use the evaluation findings.
- The local and national context within which the equity-focused programme is implemented and within which the evaluation will be conducted and used. Inequity is affected by a wide range of economic, political, social, legal and environmental factors, and the effects of all of these on the implementation, outcomes and sustainability of the interventions must be understood.

- The geographic level of the intervention (community, district, provincial, national or multi-country).
- The scale of the intervention (small, medium or large).
- The size of the evaluation budget.
- When the evaluation is commissioned (at the start, middle or end of the programme or ex-post).
- The duration of the evaluation.

It is also important to understand the methodological dimensions. While the evaluation design will be determined in part by the factors discussed above, the methodological preferences of the stakeholders of the evaluation must also be taken into consideration. At least 3 sets of factors must be considered by the evaluator:

- The required level of statistical and methodological rigor.
- The preference for quantitative, qualitative or mixed-method designs. Some stakeholders have strong feelings on the choice of an evaluation paradigm. These considerations are important for Equity-focused evaluations as many of the conventional QUANT evaluation designs used to evaluate development programmes are not adequate to capture the complex patterns of behavioral change that equity-focused interventions seek to promote. On the other hand, many groups that work on social programmes for the disadvantaged children and women believe that only qualitative methods should be used in the evaluation (making it difficult to select representative samples and to generalize from the findings).
- Whether the main source of data will be secondary, primary or a combination of both.

7.2 Reconstructing the programme theory when it is non-existent or very weak

Ideally the Equity-focused evaluation will be based on an equity-focused programme theory that was defined in a participatory way at the start of the intervention. However, evaluators will frequently be required to design an Equity-focused evaluation where:

- there is no programme theory;

- the programme theory was developed mainly by consultants, with little consultation with stakeholders; or
- there is a programme theory but it does not address equity issues.

In all of these cases the evaluators must try to “reconstruct” the implicit programme theory that explains the objectives, implementation strategy and intended outcomes of the equity dimensions of the intervention that is being evaluated.

There are at least three ways for the evaluators to reconstruct a programme’s theory:

- The *strategic approach* identifies, through group discussions with key stakeholders, the means through which the programme is expected to achieve its goals. This approach is based on a synthesis of how key actors think the programme does, or should operate, and what they think it is intended to achieve.
- In an *elicitation approach* the implicit programme theory is identified by a review of strategic documents, consultation with managers, and the observation of decision-making processes (Leeuw 2003). Field studies can also provide information which can be used to construct programme theories with, for example, the evaluator observing how the programme is explained to clients and other stakeholders by programme staff, and whether staff members encourage or discourage different groups of potential beneficiaries. This is an inductive approach that seeks to define the implicit theoretical model based on an observation of what the programme actually does. This may lead to a programme theory that differs from that based on what the actors think they do. For example, staff may believe (or at least claim) that they adopt an equity-focused approach that seeks to provide equal access to all sectors of the target population. However, observation of the programme in action may suggest that this does not actually happen.
- A *conceptualization facilitation* approach (Chen 2005) draws on the views of programme planners and stakeholders, who often have plenty of ideas about the rationale of their programme, but often do not know how to clarify their thoughts and to connect them systematically. An evaluator may facilitate this process by helping them either through forward reasoning (working from a prospective intervention to predicting its outcomes), or backward reasoning (starting from the desired outcomes and working backward to identify determinants and intervening factors). In

intensive interviews or working groups, they may identify the problem, target population, final goals and measurable outcomes, and the critical influences on outcomes. Backward reasoning may permit greater flexibility, but whether or not the group has already decided on the programme's intervention may determine whether forward or backward reasoning is appropriate.

A useful reality test is to compare the programmes assumed theory of change, derived from any of the three approaches discussed above, with information on what the programme actually does, such as (Weiss 2002):

- How funds have been allocated. If people talk a lot about the importance of something but no funds have been allocated, this is an indication that the programme component or process is not a high priority.
- The topics on which information is and is not available. A lack of available information often (but not always) suggests an aspect that is not a high priority.
- What staff members actually do. How people spend their time is another good indicator of priorities.

7.3 Conducting credible Equity-focused evaluations when working under budget and time constraints

Bamberger, Rugh and Mabry (2006 and 2012), identify the following strategies for strengthening evaluation designs when working under budget constraints:

- Simplify the evaluation design by cutting out one or more of the data collection points, for example, eliminating baseline data for the project or comparison groups, or for both groups. While this can significantly reduce the costs of data collection, the design is potentially weaker and the threats to validity increase. There are ways to strengthen the design by using secondary data (if available) as a comparison group or baseline data.
- Simplify the information to be collected. Often by eliminating non-essential information the length of the survey component can be reduced thereby saving money and time. An important consideration when identifying what information is essential and what can be discarded is to understand what key stakeholders consider to be credible evidence. While some stakeholders only

consider findings from large scale sample surveys or statistical impact evaluation designs to be credible, others accept findings from case studies, observation or focus groups. Understanding different perceptions of credibility can significantly affect the amounts and types of information that it is essential to collect.

- More economical ways to collect data. Sometimes it is possible to hire cheaper but adequately qualified data collectors (for example using nurses or teachers instead of commercial interviewers). Sometimes direct observation can replace the need for sample surveys (for example, observing what means people use to travel to work, instead of conducting a survey). Another option is to use group interviews such as focus groups or participatory rapid appraisal (PRA) techniques rather than individual interviews. Modern data collection and analysis technology such as inputting data through cell phones, hand-held computers, GPS mapping, and internet surveys, can all reduce data costs (see Box 7).
- Sometimes the cost of data collection can be shared with other agencies. For example, another agency may be willing to include a few additional questions in a planned survey; a special module could be administered to a sub-sample of households covered by another survey; or, the sampling frame could be used to identify households with certain characteristics (such as inequity criteria) to be interviewed in the equity survey.

Many of these strategies can also be used to address time constraints (for example, reducing the amount of data to be collected will also reduce time). Other ways to reduce time can be to increase the size of the data collection and analysis team or to recruit more experienced (but more expensive) researchers. Video-conferencing is another important way to save time and money, particularly during the planning and analysis stages.

Box 7. Using modern data collection technology to reduce the costs of collection and analysis of survey data

- Cell phones can reduce interview costs in a number of ways:
 - Sometimes respondents can be given cell phones so that they can be interviewed by phone, reducing the travel time and cost of interviewers. This can be useful if respondents need to be interviewed in a particular location (such as whilst using public transport, whilst transporting water or fuel on foot or queuing for water, or in the market). This also has the advantage of avoiding the need for interviewers to travel to high security risk areas.

- Responses can be recorded or directly input into the database via the phone.
- When respondents are requested to keep a diary or record of their daily activities, the information can be recorded via the phone instead of having to be written down and collected by the researcher.
- GPS-enabled phones can also be used to construct Geographical Information Systems (GIS) maps indicating the location of houses, shops, bars or other locations (see discussion of GIS below)
- Survey or observational data can be directly input on a hand-held device and possibly transmitted automatically to the central database. Analysis can also be conducted automatically.
- GIS maps are becoming increasingly available. These can indicate the location of houses, stores, public service facilities, location of road accidents, of crime or gang activity. These can sometimes be used to construct baseline data. Electronic maps are often available free or at a reasonable cost. GPS-enabled cell phones can also be used to create GIS maps as the GIS coordinates of interviews or the features such as houses, stores etc. can be automatically recorded on a GIS map.
- Internet surveys are a very economical way to collect and analyze survey data. Programmes are also available for the use of concept mapping and other more advanced forms of analysis.
- Video-cameras can be used as part of observation studies or focus group interviews. Software now makes it possible to edit video and audio-material.

7.4 Reconstructing baseline data when the evaluation is not commissioned until late in the implementation cycle²⁵

Evaluators frequently do not have access to baseline data. Not having this data significantly complicates the estimation of project impacts. There are five main scenarios under which baseline data is not available for Equity-focused evaluations:

- The evaluation was not commissioned until the project had been operating for some time.
- A baseline study was planned but never conducted.

25 For a review of strategies for reconstructing baselines see Bamberger (2010), *Reconstructing baseline data for impact evaluation and results measurement* and Bamberger (2009), *Strengthening the evaluation of development effectiveness through reconstructing baseline data*, *Journal of Development Effectiveness*. Volume 1 No. 1 March 2009.

- A baseline was conducted but does not include the information on worst-off groups and other indicators required for Equity-focused evaluations.
- The quality of the baseline design or data was too poor for the data to be used.
- The programme management was not willing to approve the conducting of the baseline study. Sometimes management will permit a baseline for the project group but not for a comparison group and in other cases they will not authorize any baseline study.

There are a number of strategies that can be used to “reconstruct” baseline data, all of which apply to Equity-focused evaluations:

- Using data from the programme M&E system and the administrative records.
- Using records from other organizations (schools, health clinics etc.) to construct a comparison group.
- Using records from national data sets (MICS, LSMS etc.).
- Using recall: respondents are asked to recall their situation or that of their community or group at the time the project began. For example, respondents can be asked to recall their income or expenditures; travel time to work, or to collect water or fuel; which children attended school outside the village before the project school was built, etc. While recall is often the only available source of information on the past, the challenge is that it is difficult to detect potential sources of bias (from problems with memory, difficulties of locating events in time or sometimes intentional distortion). In most cases there are also no guidelines to estimate and adjust for the direction and magnitude of bias.
- Key informants.
- Focus groups.
- PRA and group consultation techniques.

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 of the more subtle and sensitive concepts relating to equity than it is to obtain relatively straightforward information on things like school enrolment and travel time. PRA tech-

niques have been used quite extensively to identify social categories using techniques like wealth ranking and social mapping. When working with small communities such as a village, it has been possible to construct a social map of the community, and to rank every household in terms of wealth or sometimes other vulnerability characteristics.

7.5 Conducting Equity-focused evaluations in countries where government is not supportive of an equity focus

Section 3 reviewed the challenges in introducing Equity-focused evaluation approaches under different scenarios. Many of the issues refer to scenarios where national governments or particular agencies are not supportive, or may actively oppose equity approaches in general, or the introduction of Equity-focused evaluation in particular. The opposition may be due to lack of support for worst-off groups or the desire to discourage them from entering the country or moving to particular areas; to reluctance to change the indicators the agency currently uses to assess progress on poverty and social development; to the limited capacity to conduct more complex evaluations; or, to cover the extra costs of conducting these evaluations. UNICEF may also have the problem of not being able to provide, directly or through partner agencies, the additional financing that might be required for these studies.

These issues are difficult to address as they often combine political, financial and technical/capacity questions. The following are some of the strategies that can be considered, several of which use the real-world evaluation strategies discussed in this section:

- Try to reduce the additional costs of data collection (for example for the administration of special modules) by combining with surveys planned by government, civil society or other donors.
- Coordinate with sector programmes to identify areas where equity interventions are likely to be relatively non-controversial and would produce some immediate gains. For example, providing transport might dramatically increase the number of low-income mothers who bring their children for health check-ups. The evaluations would then demonstrate the benefits of these interventions.

- A similar approach could be used to identify programmes where Equity-focused evaluation could identify cost-effective interventions to increase accessibility in non-controversial ways. Equity-focused evaluations can often use the bottleneck supply and demand framework to identify potential areas of intervention. The analysis of demand-side factors can identify some constraints on use of services that can easily be addressed. A good starting point is to examine factors affecting the ability of low-income women, and particularly mothers, to access services. Often this will identify issues such as cost, lack of transport, and inconvenient opening hours or locations, many of which could be addressed relatively easily.
- Develop simple guidelines and checklists that the evaluation departments of government agencies could use for collecting equity-related information.

SECTION 8: CASE STUDIES OF UNICEF-SUPPORTED EQUITY-FOCUSED EVALUATIONS

The following case studies 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 targeted 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 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.

REFERENCES

- Bamberger, M. (2009). *Strengthening the evaluation of development effectiveness through reconstructing baseline data*. Journal of Development Effectiveness. Volume 1 No. 1 March 2009.
- 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. & Ooi, E (Eds) (2005). *Influential Evaluations: detailed case studies*. Case study No. 3 *Using Citizen Report Cards to Hold the State to Account in Bangalore, India*. Operations Evaluation Department. The World Bank. Available at: www.worldbank.org/oed/ecd.
- Bamberger M. & Ooi, E (Eds) (2005). *Influential Evaluations: Detailed case studies*. Case study No.7 *Improving the delivery of primary education services in Uganda through public expenditure tracking surveys*. Independent Evaluation Group. The World Bank. Available at www.worldbank.org/oed/ecd.
- Bamberger, M., Rao, V., & 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 613-642.
- Bamberger, M. Rugh, J. & Mabry, L. (2006). *Real World Evaluation*. Sage Publications.
- Bamberger, M., Rugh, J. & Mabry, L. (2012). *RealWorld Evaluation Second Edition*. Sage Publications.
- Chen, H. (2005) *Practical Program Evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage Publications.
- Convention on the Rights of the Child (CRC) (1990).
- Davoodi, H., Tiongson, E. & Asawanuchit, S. (2003). *How useful are benefit incidence analyses of public education and health spending?* IMF Working Paper. International Monetary Fund.
- DFID (2008). *Reducing poverty by tackling social exclusion: A DFID policy paper*.
- Donaldson, S., Christie, C. & Mark, M. (2009). *What counts as credible evidence in applied research and evaluation practice?* Sage Publications.
- Eckman, K. & Walker, R. (2008). *Knowledge, attitudes and practice (KAP) survey: summary report from the Duluth Lakeside Stormwater Reduction Project*. Water Resources Center. University of Minnesota. (Downloaded from "Google").
- Fluke; J. & Wulczyn, F. (2010). *A concept note on children protection systems*.
- Funnell, S. & Rogers, P. (2011). *Purposive programme theory: effective use of theories of change and logic models*. Jossey-Bass Publications.
- Gertler, P. et al. (2011). *Impact Evaluation in Practice*. The World Bank.

- 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 n°. 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.
- Kaliyaperumal, K. (2004). *Guidelines for conducting knowledge, attitude and practice (KAP) study*. Community Ophthalmology Vol IV. No. 1 Jan-Mar 2004. (Downloaded from "Google").
- Khandker, S., Koolwal, G. & Samad, H. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. The World Bank.
- Levin, H. & McEwan, P. (2001). *Cost-effectiveness Analysis: Methods and Applications*. Sage Publications.
- Levin, H. (2005). *Cost effectiveness*. In Sandra Mathison (editor) *Encyclopedia of Evaluation*. Sage Publications.
- Mackay, K. (2007). *How to build M&E systems to support better government*. Independent Evaluation Group. The World Bank.
- Mayne, J. (2008) "Contribution analysis: An approach to exploring cause and effect." Rome: Institutional learning and change initiative , ILAC Brief No. 16. May 2008. http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf
- Muennig, P. (2008) *Cost-effectiveness analysis in health: A practical approach*. Wiley Publications
- 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.
- Organization for Economic Cooperation and Development. 2010 (Second Edition). *Evaluating Development Cooperation: Summary of Key Norms and Standards*. Paris: OECD. <http://www.oecd.org/dataoecd/12/56/41612905.pdf>
- Paes de Barros, R., Henriques, R., & Mendonça R. (2002). *Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil*. Texto para discussão n°. 857. Rio de Janeiro: IPEA.
- Patton, M.P. (2011). *Developmental Evaluation: applying complexity concepts to enhance innovation and use*. Guilford Press.
- Picciotto, R. (2002). *International trends and development evaluation: the need for ideas*. *American Journal of Evaluation*. 24 (2): 227-34
- Pickett, 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.
- Rogers, E. (2003). *Diffusion of Innovations*. Free Press.
- Rossi, P., Lipsey, M. & Freeman, H (2004). *Evaluation: A systematic approach*. Sage Publications.
- 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 n° 9*. Florencia, Italia: UNICEF Innocent Research Centre.
- Segone, M. (1998). *Democratic Evaluation: A proposal for strengthening the evaluation function in International Development Organizations*.
- 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 monitoring and evaluation systems*. UNICEF, Devinfo, IDEAS, IOCE, MICS, UNECE and The World Bank.
- 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>
- Swedish International Development agency (SIDA) (1999). *Are Evaluations Useful: Cases from Swedish Development Cooperation*. Department for Evaluation and Internal Audit.
- Tanahashi, T. (1978). *Health service coverage and its evaluation*. Bulletin of the World Health Organization. 56 (2): 295-303.
- UNDP. Human development Report (2010). *The real wealth of nations: Pathways to Human Development*.
- UNEG (2005). *Standards for Evaluation in the UN System*
- UNEG (2011). *Integrating human rights and gender equality in evaluation*.
- UNICEF (2003). *Report on the Situation of Children and Adolescents in Brazil*.

- UNICEF (2008). *Child Protection MetaEvaluation*. Commissioned by Child Protection Section, Programme Division.
- UNICEF (2009a). *Children and the 2004 Indian Ocean Tsunami: Evaluation of UNICEF's response in Indonesia, Sri Lanka and Maldives, 2005-08*. Overall synthesis report.
- UNICEF (2009b). *Roundtable meeting on child protection Monitoring and Evaluation: Progress, challenges and the way forward*: Summary Report. With Save the Children.
- 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 (2010e). *Better Care Network. Manual for the measurement of indicators for children in formal care*.
- UNICEF (2010f). Ager, A., Akesson, B & Schunk, K. *Mapping of child protection M&E tools*.
- UNICEF (2010g). *Advocacy toolkit. A guide to influencing decisions that improve children's lives*.
- 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.
- UNICEF, USAID, Oak Foundation, World vision, Save the Children (2009). *What are we learning about protecting children in the community: an interagency review of evidence on community based child protection mechanisms*.
- United Nations News Centre (2001). *UN and Oxford University unveil new index to measure poverty*, <http://www.un.org/apps/news/story>.
- United Way of America (1999). *Community status reports and targeted community interventions: drawing a distinction*. Alexandria, VA. United Way of America.
- Victoria, C., Black, R. & Bryce, J. (2009). *Evaluating child survival programmes*. Bulletin of the World Health Organization. 2009:87:83.
- Weiss 2001. Theory based evaluation: Theories of change for poverty reduction programs pp. 103-14 in *Evaluation and Poverty Reduction* edited by O. Feinstein and R. Piccitto. New Brunswick. Transaction
- Williams B. (2005). *Systems and Systems Thinking* in Mathison, S (editor) Sage Encyclopedia of Evaluation pp. 405-412.
- Williams, B. (2010). *Systems Thinking and Capacity Development in the International Arena*. In: Fujita (ed), *Beyond LogFrame. Using systems concepts in evaluation*, DASID, Japan.
- Williams, B. & Imam I. (Eds.), (2007). *Systems Concepts in Evaluation: An expert anthology*. American Evaluation Association.
- Wholey, J., Hatry, H. & Newcomer, K. (2010). *Handbook of Practical Programme Evaluation*. Jossey-Bass.

UNICEF Evaluation Office
3 United Nations Plaza
New York, NY 10017, USA
<http://www.unicef.org/evaluation/index.html>

2011