UN WOMEN

EvalGender+

ioce

Evaluating the Sustainable Development Goals

With a "No one left behind" lens through equity-focused and gender-responsive evaluations

In partnership with:

EvalPartners

UNEG
United Nations Evaluation Group

unicef

UNFPA

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

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

ISSUE # 2: Bridging the gap: The role of monitoring and evaluation in evidence-based policymaking. Published by UNICEF in partnership with 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, World Bank, UNDP, UNIFEM, and WFP, 2010.

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

ISSUE # 7: Evaluation for equitable development results. Published by UNICEF in partnership with Coneval, IDEAS, IDRC, ILO, IOCE, UNDP, and UN Women, 2012.

ISSUE # 8: Evaluation and civil society: Stakeholders' perspectives on national evaluation capacity development. Published by UNICEF, EvalPartners and IOCE in partnership with CLEAR, IEG World Bank, Ministry for Foreign Affairs of Finland, OECD Development Assistance Committee Network on Development Evaluation, UNEG, and UN Women, 2013.

ISSUE # 9: Voluntary organizations for professional evaluation: Learning from Africa, Americas, Asia, Australasia, Europe and Middle East. Published by UNICEF, EvalPartners and IOCE in partnership with Cooperación Española, Ministry for Foreign Affairs of Finland, UNEG, and UN Women, 2013.

ISSUE # 10: Advocating for evaluation: A toolkit to develop advocacy strategies to strengthen an enabling environment for evaluation. Published by UN Women, EvalPartners, IOCE in partnership with Ministry for Foreign Affairs of Finland, OECD, UNEG, UNICEF, and USAID, 2014.

ISSUE #11: National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems. Published by UN Women, EvalPartners, IOCE in partnership with USAID, UNEG, Ministry for Foreign Affairs of Finland and Parliamentarian Forum for Development Evaluation, 2015.

Photo Credits: ©UN Photo

All the above publications are available for free download at www.evalpartners.org/selected-books.

Disclaimer:

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

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

Evaluating the Sustainable Development Goals

With a "No one left behind" lens through equity-focused and gender-responsive evaluations

Authors
Michael Bamberger, Marco Segone and Florencia Tateossian

Foreword

We have entered the exciting era of the full implementation of the 2030 Agenda for Sustainable Development. This Agenda comes together with a follow-up and review mechanism to ensure the Sustainable Development Goals (SDGs) are systematically monitored and reviewed to help countries implementing the 2030 Agenda to ensure "No one is left behind". The UN General Assembly underlined the importance of evaluation within the transformative 2030 Agenda calling for: (a) the review and follow-up mechanisms to be informed by country-led evaluations, and (b) capacity-building support for developing countries including strengthening of national data systems and evaluation.

This means evaluation should play a crucial role to support effective and efficient SDG implementation. Evaluation will offer evidence-based learning on how policies and programmes delivered results and what needs to be done differently.

But this is not enough. The main principle of the 2030 Agenda is that no one should be left behind. The follow-up and review mechanisms also call for inclusiveness, participation and ownership. This is why equity-focused and gender-responsive evaluation is needed. This transformative kind of evaluation can help countries to identify structural causes of inequalities through deeper analysis of power relationships, social norms and cultural beliefs. Integrating EFGR evaluations will provide strong evidence to ensure national voluntary reviews of SDGs are leaving no one behind.

Marco Segone

Co-Chair, EvalGender+
Director, Independent Evaluation Office, UN Women
Chair, United Nations Evaluation Group (UNEG)
marco.segone@unwomen.org
@msegone

Contents

Acknowledgements **7**

Acronyms **8**

Introduction **9**

1. The 2030 Agenda for Sustainable Development: No one left behind .. **11**

1.1 The origin and overarching goals of the SDGs 11

1.2 The 2030 Agenda is a political agenda 14

1.3 The key differences between the MDGs and the SDGs 15

1.4 Implications of the data revolution for the evaluation of the SDGs 16

1.5 Rethinking evaluation methodology 18

1.6 The implications of gender and reducing inequalities dimensions of the SDGs 26

1.7 Understanding the respective responsibilities and challenges for less developed, middle income and developed countries in the implementation of the evaluations 27

1.8 Tailoring the evaluations of SDGs to national, sub-national and local contexts 28

2. SDG follow-up and review mechanisms **30**

2.1 The SDG framework for follow-up and review 30

2.2 The role of different actors at the national and subnational levels 30

2.3 Strengthening and using existing M&E systems at the local, national and international levels 32

2.4 Steps towards strengthening the capacity of existing M&E systems to conduct gender analysis and to address inequality issues 34

2.5 Assessing the quality and coverage of data on gender equality and data on reducing inequalities in the SDGs 36

3. Relevant principles of the 2030 Agenda for a "No one left behind" evaluation **40**

3.1 Framing evaluation within the SDG principles 40

3.2 Gender equality, reducing inequalities and ensuring "No one left behind" 40

3.3 Sustainable development: Sustainability and resilience 42

3.4 Interlinkages across SDGs: Complexity and complex development programmes 44

4. Proposed equity-focused and gender-responsive evaluation framework for SDGs 57

4.1 The proposed country-level, equity-focused and gender-responsive evaluation framework 58

4.2 Main types of evaluation designs that can be considered for equity-focused and gender-responsive evaluations 65

4.3 Some general strategies for integrating equity-focused and gender-responsive principles into SDG evaluations 72

4.4 Approaches for the evaluation of complex development programmes 75

4.5 Evaluating sustainability and resilience in the SDGs 78

5. Proposed strategy to integrate equity-focused and gender-responsive evaluations into national SDG reviews 80

5.1 Integrating an equity-focused and gender-responsive lens into the national SDG evaluation policies and systems 80

5.2 Identifying stakeholders concerned with gender and reducing inequalities 88

5.3 Developing an equity-focused and gender-responsive national evaluation system 89

5.4 Developing an advocacy strategy to promote equity-focused and gender-responsive evaluation use 89

Annex 1. Lessons from the MDGs 93

Annex 2. Potential applications of big data and new information technologies for EFGR evaluations 99

Annex 3. Feedback from online consultations on the design of equity-focused and gender-responsive evaluations for the SDGs 101

Acknowledgements

The editors and authors would like to acknowledge and thank the contributions of several colleagues, including the **members of EvalGender+ Management Group** (composed of Awuor Ponge, IDEAS; Fabiola Amariles, REDWIN; Maha Al Said, EVALMENA; Adeline Sibanda, African Evaluation Association; Tatiana Tretiakova, Eurasia EvalSociety; Rashmi Agrawal, Community of Evaluators India; Svetlana Negroustoueva, American Evaluation Association; Luisa Belli, FAO Evaluation Office; Alejandra Faundez, ReLAC; Alexis Salvador Loye, Réseau Francophone de l 'Evaluation; Madri Jansen van Rensburg, AGDEN; Julia Espinosa Fajardo, European Evaluation Association; Ada Ocampo, UNICEF Evaluation Office; Rituu B. Nanda, Gender and Evaluation Online Community of Practice; Isha Wedasinghe Miranda, Asia-Pacific Evaluation Association; Hon Susan Musyoka, Global Parliamentarian Forum for Development Evaluation; Sonal Zaveri, Community of Evaluators South Asia; Asela Kalugampitiya, EvalPartners Coordinator; and Keiko Kuji-Shikatani, Canadian Evaluation Society); the **Independent Evaluation Office of UN Women**; the **Research and Data Section of UN Women**, in particular Ginette Azcona, Somali Cerise, Sophie Brown and colleagues from **other UN agencies** (Colin Kirk, UNICEF Evaluation Office; Ada Ocampo, UNICEF Evaluation Office; Andrea Cook, UNFPA Evaluation Office; Alexandra Chambel, UNFPA Evaluation Office; and Natalie Raaber, UNFPA Evaluation Office).

The publication was enriched by debates during the "No one left behind" event held in New York, 15-17 March 2016. More information is available at: http://evalpartners.org/evalgender/evaluating-SDGs-with-an-equity-focused-and-gender-responsive-lens.

Acronyms

EFGR	Equity-focused and Gender-responsive
IAEG-SDGs	Inter-Agency Expert Group on SDGs
M&E	Monitoring and Evaluation
MDG	Millennium Development Goal
NEP	National Evaluation Policy
QCA	Qualitative Comparative Analysis
RCT	Randomized Control Trial
SDG	Sustainable Development Goal
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
UN Women	United Nations Entity for Gender Equality and the Empowerment of Women

Introduction

The purpose of this publication is to provide guidance on how to integrate an equity-focused and gender-responsive (EFGR) approach to national evaluation systems that should inform national Sustainable Development Goal (SDG) reviews. The 2030 Agenda for Sustainable Development puts forward "a plan for action for people, planet and prosperity" and "seeks to strengthen universal peace in larger freedom" through strategic partnerships. It includes a vision and principles, a results framework of global SDGs, a framework for means of implementation and global partnership, and a follow-up and review mechanism.

This guidance is intended to support national evaluation systems on how to integrate EFGR evaluations to inform the national reviews of SDGs. The purpose of this document is: (a) to provide guidance on how to integrate an equity-focused and gender responsive approach to national evaluation systems generally, and (b) to propose a step by step process for country-led evaluations that are equity-focused and gender-responsive as well as a strategy to integrate equity-focused and gender-responsive evaluations to inform national SDG reviews. Gender statistical analysis and disaggregated indicators, while essential, will need to be complemented by EFGR evaluation approaches to ensure no one is left behind.

The guidance is expected to primarily serve national evaluation systems, the UN system, multilateral and bilateral development agencies, academic institutions, including specialized research centres and think tanks, private foundations, the private sector, and voluntary organizations of professional evaluators.

The publication is organized in five chapters. Chapter 1 presents the SDG agenda and the central principle of ensuring "No one is left behind". The chapter begins by discussing the origin and overarching goals of the SDGs, stressing that the 2030 Agenda is a political agenda and highlights principles related to gender equality, human rights, leaving no one behind, universality and interconnectedness of various SDG goals. This chapter introduces the key differences between the Millennium Development Goals (MDGs) and the SDGs, including from an evaluation perspective. The chapter also discusses the implications of the data revolution for the SDG evaluations and puts forwards the major advances in evaluation methodology to be used in the evaluation of SDGs.

Chapter 2 presents the proposed SDG framework for follow-up and review. While the current framework does not focus directly on the

proposed evaluation methodology, it lays out the principles—emphasizing that the approach is voluntary and country-led, and recognizing the different national realities, capacities and levels of development. The central principle is to ensure "No one is left behind."

Chapter 3 discusses the relevant principles of the SDG agenda with the lens of "No one left behind", notably: (a) gender equality and reduction of inequality and how it applies to equity-focused and gender-responsive evaluation, (b) sustainable development and how it applies to evaluation, and (c) interlinkages across the SDGs and the need for complexity-responsive evaluation.

Chapter 4 proposes a framework for the design and implementation of EFGR evaluations at the country level. The 2030 Agenda proposes a set of standard indicators and sub-indicators for measuring each of the SDGs, which all countries are requested to collect and analyse, as far as their capacities permit, using the standard definitions. Countries are also encouraged to conduct further monitoring and evaluation (M&E) studies that respond to their particular priorities and concerns with respect to the SDGs. This chapter proposes a framework that countries can use for developing their own country-led evaluations that are EFGR. The framework presents a nine-step strategy for a comprehensive EFGR evaluation strategy at the country level.

Last but not least, chapter 5 proposes a strategy to integrate EFGR evaluation into national evaluation systems and SDG reviews. The political and methodological challenges of integrating an EFGR approach are discussed. Based on lessons from past experience, a set of guidelines are proposed for ensuring the effective integration of EFGR approaches.

1. The 2030 Agenda for Sustainable Development: No one left behind

The guidance is based on two concepts directly derived from the SDGs, specifically:

- Gender equality: Derived from SDG 5
- Reduce inequality within and among countries: Derived from SDG 10

These two denominations are summarized throughout the document as gender and reducing inequalities.

When referring to evaluation approaches for addressing these, the term equity-focused and gender-responsive (EFGR) evaluation is used.

1.1 The origin and overarching goals of the SDGs

A new world vision—The world we want

The SDGs, which were approved by the UN General Assembly in 2015, seek to build on the MDGs, approved in 2000, and to "complete what they did not achieve".[1] Consequently, the SDGs will continue the momentum generated by the MDGs and seek to learn from this ongoing international commitment—both with respect to what worked and what did not.

While most of the major themes of the MDGs continue to be reflected in the SDGs, the SDGs are broader and more ambitious. While the SDGs continue to focus on poverty alleviation and prosperity, they also reflect the growing concerns about the future of our planet as reflected in the focus on sustainability and the need for social inclusion to ensure "No one is left behind."

The SDGs also present a vision of "the world we want". This combines issues of climate change, environmental conservation, strengthening community organization and broad-based participation in the political process. Sustainability also concerns peace and security and the strengthening of partnerships among all of the actors in the development process. All of these dimensions are reflected in the five areas of critical importance for the SDGs: people, planet, prosperity, peace and partnership—all of which are interlinked in complex and evolving ways.

1 United Nations. 2015. "Transforming Our World: The 2030 Agenda for Sustainable Development".

The principle of universality

"Sustainable development is a universal aspiration requiring international cooperation and joint responsibility for a greater common good."[2] In 2012, the Rio + 20 Outcome Document called for the SDGs to be "global in nature and universally applicable to all countries while taking into account different national realities and levels of development."[3]

The fact that the SDGs are based on five universal and interlinked principles—people, planet, prosperity, peace and partnership—means that the evaluation must have a broader, holistic and more complex approach than for the MDGs. The UN Environment Programme and UN High Commissioner for Refugees identified four aspects of the notion of universality[4]:

- The recognition of universal principles, standards and values applicable to all countries and all peoples;
- The interconnectedness of national and global challenges, and therefore universal commitments to address them;
- Recognizing that sustainable development issues exist in all countries;
- A universal commitment to "No one left behind";

As will be discussed later, the principles of universality have several implications for the design of an evaluation of SDGs:

- Universality implies connectedness, which means that it is important to recognize that the outcomes of each goal are closely affected by the outcomes of all other goals;
- This in turn requires the use of a complexity-responsive evaluation design;
- A human rights focus must inform both the implementation of an evaluation and also the evaluation process;
- An EFGR evaluation is also strongly required to ensure no one is left behind;
- Finally, the commitment to "No one left behind" requires a focus on equity, voice and empowerment as well as close attention to unintended outcomes of development interventions, which often result

2 UN Environmental Programme. Undated. "Universality in the post 2015 Sustainable Development Agenda". Post 2015 Note #9 / OHCHR Human Rights and Post 2015. pp. 1-2 (undated). Available at: http://www.ohchr.org/Documents/Issues/MDGs/Post2015/OHCHR_UNEP.pdf

3 Ibid.

4 Ibid.

in women and other vulnerable people being excluded from equal access to the benefits of many development interventions as well as public policies and national processes.

The challenge of the country-driven focus

A key element of the SDG approach is that it is country-led and participatory. The programmes and the evaluations should be owned and led by a wide range of government, civil society and community organizations. While offering new opportunities to draw on the resources and to give voice to a much broader range of organizations and groups, the approach also brings new challenges. One challenge involves the coordination among so many different organizations with different perspectives and priorities and different approaches to evaluation. There may also be different approaches to evaluation, with some agencies traditionally using quantitative methods while others use qualitative and participatory methods.

Prioritizing gender equality and women's empowerment

While gender equality was considered somewhat important in the MDGs, it was assumed that all of the issues could be captured in MDG-3 (promote gender equality and empower women) and the framework did not directly address cross-cutting issues, such as the effects of infrastructure or agriculture on women (although these connections were not completely ignored). In contrast, the SDGs recognize and address the inter-linkages between gender and the other SDGs, and at least 10 of the SDGs include an indicator referring to women's equality. The SDGs also focus more on empowerment and equality.

Finally, the central goal of "No one left behind" also recognizes the gender dimensions of exclusion.

The focus on social inclusion—"No one left behind": gender and equality

The 2030 Agenda proposes a social inclusion approach where the goal is to ensure that no one is left behind. This is much more challenging—technically and politically—than conventional approaches where the goal is simply to increase the proportion of the population with access to services and programmes. While there is usually broad consensus on goals such as increasing school enrolment, improving road networks or access to water supply and sanitation, there is frequently less agreement on how more resources and effort should be devoted

to reaching the poorest and most vulnerable groups, including women and girls, ethnic and religious minorities, refugees, illegal immigrants and people differently abled.

Social inclusion has a strong political dimension. In many societies, there is strong social disapproval of some minorities or vulnerable groups, and there may be little support for trying to integrate these groups. Also, a focus on the identification of groups who have been excluded will require governments to acknowledge that their social and economic performance has not been as strong or effective as their ranking on key international indicators, such as the Human Development Index, would suggest.

There are also logistical and methodological challenges. Comparative data is usually not available on the access of different religious or ethnic groups to basic services. Consequently, a social inclusion focus will require collection of additional data and new methodologies for data collection and analysis. The additional costs and need for greater evaluation expertise are also major disincentives for agencies working under budget and time constraints.

1.2 The 2030 Agenda is a political agenda

The 2030 Agenda presents a bold vision of a transformed world with a much greater commitment to social and political justice, a recognition of the need to adapt development strategies to the constraints of the planet, new forms of social and political accountability, and a focus that seeks to include marginal and vulnerable groups. A focus on the bottom 40 per cent will require courageous political commitments to reverse the trend towards the increasing concentration of wealth and decision-making power in the hands of an ever smaller fraction of the population. This requires a transformative approach, which may meet with strong opposition from those in power or sectors that have previously been largely unregulated.

The 2030 Agenda also has a strong value orientation based on social justice. The role of values will be particularly important in determining which excluded groups will be included. This will also be a highly sensitive political process.

The approach recognizes that inequalities are structural and result from unjust political and social systems and that addressing these will require a dramatic and difficult transformation. Giving voice to sectors that have previously been unheard and excluded will require fundamental social

and economic rethinking and reorganization. The transformation requires support at both the highest and grass-roots levels.

Although seeming to be a purely technical issue, the emphasis on an evidence-based decision-making process also has strong political implications. It means that evaluation is built into the political decision-making process and challenges programmes that are not based on empirical evidence on their efficacy.

Finally, the fact that the evaluation process will be country driven and participatory is also a political decision. On one level, it is intended to ensure that the SDGs are "owned" by the countries and not by the United Nations and donor agencies; while on another level, it means that a much broader range of country and local-level stakeholders are involved including civil society organizations.

1.3 The key differences between the MDGs and the SDGs[5]

While the SDGs build on the MDGs and are intended to take over from where the MDGs left off, there are some important differences between the two, including:

- As discussed in Section 1.1, the SDGs introduce the principle of universality, which was not a part of the MDGs.
- While the MDGs focus on the extent to which targets have been achieved, the SDGs also assess the extent to which the outputs and outcomes are sustainable over time.
- While the MDG evaluation conducted independent assessments of each of the eight MDGs, the SDGs recognize the complementarities among the individual SDGs and that the achievement of any individual SDG is dependent on the contribution of other SDGs.[6] For example, improvements in health or education are dependent on national and international economic trends and the status of poverty reduction in a particular country, and on the availability of roads, water, energy and transportation, among others.
- There are a number of cross-cutting SDGs, such as gender equality and reducing inequalities, that are assessed in their own right but that are also key determinants of the success of most other SDGs.

5 See Annex 1 for a summary of lessons from the evaluation of the MDGs.

6 Le Blanc, D. 2015. "Towards integration at last? The sustainable development goals as a network of targets". DESA working paper No. 141, March 2015.

- In terms of addressing gender equality, another limitation of the MDGs was the narrow focus of the targets, which detracted resources and attention from the root causes of gender inequality addressed in the more comprehensive normative agreements on gender equality, such as the Convention on the Elimination of All Forms of Discrimination against Women and the Beijing Platform for Action.[7]
- While the MDGs were largely assessed in terms of aggregate indicators, such as the increase in the proportion of individuals, households and communities achieving a certain target (e.g., the proportion of girls enrolled in school or households with access to safe water), the proposed framework for an evaluation of the SDGs will, like the SDGs themselves, focus on inequality between and within countries, including a specific focus on achieving gender equality and reducing inequalities. This requires, inter alia, asking the question "Were any groups left behind?" and assessing the gap between the poorest and most vulnerable groups and the rest of the population.
- The process of developing the SDGs was done in a higher participatory and inclusive manner, involving many more groups than the MDGs.

1.4 Implications of the data revolution for the evaluation of the SDGs[8]

There has been a dramatic increase in the use of smart phones, tablets, automatic sensors and other new information technology for the collection, analysis and dissemination of data for programme design and M&E. More recently, there has been an exponential growth in the availability of big data such as the analysis of tweets and social media, analysis of electronic financial transactions, phone records and satellite images. Information can be collected much faster and cheaper, and information can be collected and disseminated in real-time so that it can provide early warnings of natural and man-made emergencies.

There have also been important advances in smart data analytics with the power to identify new patterns and relationships in the data that

7 Sen, G. and A. Mukherjee. 2014. "No empowerment without rights, no rights without politics: Gender-equality, MDGs and the post-2015 development agenda".

8 For a discussion of the applications of big data for programme evaluation, see Bamberger (to be published later in 2016) "Guidelines for integrating big data into the monitoring and evaluation of development programmes".

were previously difficult to detect.[9] New technologies for data visualization, such as interactive maps, also make it possible to present data in ways that are easily understandable to community groups and other stakeholders who are not data analytic specialists.[10]

Over the past two years, there has been a growing interest in the application of big data for gender-responsive evaluation. The Data2X collaboration is perhaps the most advanced of these initiatives (see Box 1).

BOX 1. The Data2X collaboration to promote the use of big data for gender

In collaboration with UN Global Pulse, UN Women and individual academic researchers, Data2X is spearheading research pilots to explore how different methods of collecting and analysing big data could potentially close global gender gaps. The partnership will also devise a long-term strategy for expanding the use of big data for gender within UN Global Pulse, and aims to be a springboard for future big data for development efforts to ensure that gender remains at the forefront of this nascent field. Some of the areas where pilot initiatives have been launched include:

- Civil registration and vital statistics
- Women's work and employment
- Supply side data on financial services
- Women's subjective well-being and poverty
- Big data and gender
- Improved gender data on U.S. foreign assistance programmes
- Data on displaced populations

Source: http://data2x.org/

Annex 2 presents examples of the use of big data and new information technology with potential applications for EFGR evaluations. Currently, most of the applications of these technologies are used for research, programme design and emergency relief rather than directly for programme evaluation—although many techniques could be adapted for programme evaluation.

9 For references on data analytics, see Marr, B. 2015. *Big Data: Using Smart Big Data Analytics and Metrics to Make Better Decisions and Improve Performance.* Wiley; Meier, P. 2015. *Digital Humanitarians: How Big Data is Changing the Face of Humanitarian* Response. CRC Press; and Siegel, E. 2013. *Predictive Analytics: The Power to Predict Who Will Click*, Buy, Lie or Die. Wiley.

10 For references on applications of data visualization see Meier (op.cit); World Bank, World Development Report 2016 Digital Dividends.

An important challenge for evaluations of SDGs will be to assess the potential applications of all of these new information technologies and how they can be combined with conventional evaluations. There is a tremendous potential for these new technologies, and SDGs must help evaluators to catch up and use these technologies. It will, of course, be necessary to assess the cost and accessibility of the technology. There are also important ethical and political issues to address, including data security and privacy and the fact that access to many kinds of big data is often limited to governments and well-resourced international agencies.[11] There is a concern that big data may become "extractive" in that it provides a way for governments and donor agencies to collect information without having to interact with local communities, often without these communities even being aware that data is being collected about them. While methods exist to ensure that big data promotes inclusion and participation, a strong advocacy campaign will be required to ensure that big data applications contribute to the goal of "No one left behind."

1.5 Rethinking evaluation methodology

Since the launch of the MDGs in 2000, there have been major advances in evaluation methodology that must be incorporated into the planning of evaluation of SDGs. All of these have important implications for the EFGR evaluation.

Main types of evaluation and key evaluation questions

While much of the evaluation literature focuses on evaluating the *impacts* of *development projects,* it is important for the design of SDG evaluations to recognize that a comprehensive evaluation of the SDG at national level must be conducted at three levels, and that there are at least four main evaluation approaches.

The three levels are:

- The evaluation of national and sector policies;
- The evaluation of broad-based programmes that normally involve a number of different components or projects. These tend to involve a large number of implementing agencies and stakeholders, have a broad geographical coverage, and often have not very complete information on where and how each component was implemented

11 Bamberger, M., Raftree, L., and V. Olazabal. 2016. "The role of new information and communication technologies in equity-focused evaluation: Opportunities and challenges". *Evaluation*, April 2016, 22 (2): 228-244.

- The evaluation of projects that usually, but not always, have a relatively limited geographical coverage and a limited number of components and implementing agencies.

The four most common evaluation approaches are:

- ***Policy evaluation:*** This assesses how well policies and broad programmes (such as country programmes and multi-donor collaborative programmes) are designed and implemented and how well they achieve their development objectives. These evaluations focus on upstream development and planning. Many of the evaluations are conducted retrospectively, often at the end of a country programme cycle (typically lasting four to five years). Many of these evaluations use the Organisation for Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC) evaluation criteria (e.g., relevance, efficiency, efficacy, impact and sustainability), but many other policy evaluation methodologies can be used.
- ***Formative evaluation:*** The purpose of formative evaluation is to provide regular feedback to management and other stakeholders to help strengthen the implementation of programmes and projects. There is a close linkage between monitoring and evaluation to ensure that maximum use is made of monitoring as a tool for agile management and not just for accountability. Formative evaluation combines quantitative and qualitative methods, often combined into a mixed methods approach. There is also a focus on evaluation as a learning tool. Formative evaluation is used throughout the programme and project cycle. The approach is based on close collaboration between management and the evaluation team, and the organizational approach is distinct from many summative evaluations, which often stress that "objectivity" can only be achieved by maintaining a distance between managers and evaluators. Many kinds of formative evaluation also include a rights-based approach, which employs qualitative and participatory approaches to use evaluation to give voice to poor and vulnerable groups and to promote social justice. Many forms of EFGR evaluation fall into this category.
- ***Developmental evaluation:*** In recent years, many agencies include Michael Patton's developmental evaluation[12] as a fourth type of evaluation. This has many similarities with formative evaluation in that the purpose is to help managers and other stakeholders improve programme performance and learn lessons for the selection and

12 Patton, M. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* New York, NY: Guilford Press.

design of future programmes. However, developmental evaluation focuses on innovative programmes and those that operate in complex environments where an adaptive approach to design and implementation must be used. Interventions evolve and adapt, and often they do not have any completion point. The approach is based on a very close collaboration between managers and evaluators, where the latter are closely involved in programme implementation and adaptation and not as external evaluators.

- ***Summative evaluation:*** The purpose of summative evaluation is to assess the extent to which observed changes in outcome variables (the intended project goals) can be attributed to the effects of the project intervention. These evaluations can either be quantitative, estimating the size and statistical significance of the changes, or they can adopt a more qualitative approach—where one of the main sources of evidence is the opinions of the affected populations and other stakeholders. Traditionally, summative evaluations have been used for accountability and to provide guidance on the potential replicability of programmes. The most widely used tool for summative evaluation has been randomized control trials (RCTs); RCTs are also one of the most criticized evaluation methodologies. An exclusive focus on RCTs is widely challenged within the evaluation community due, among other things, to: a narrow focus on a single or small number of (usually quantitative) outcomes; and a lack of attention to the process of project implementation and to the context within which programmes are designed, implemented and evaluated. RCTs are also challenged by rights-based evaluators who stress the need to listen to multiple voices and who argue that there is no one way to identify or assess programme outcomes. An important development is the "RCT+" approach, which combines experimental evaluation designs with qualitative approaches.[13]

Table 1 lists some of the key questions that each of the four types of evaluation must address. It is important to recognize that each type of evaluation is designed to address different kinds of questions. Consequently, it is important to identify the kinds of questions of concern to different stakeholders before selecting the evaluation design. It will often be the case that more than one type of evaluation may be required to address all of the questions of interest to stakeholders.

13 Bamberger, M., Tarsilla, M., and S. Hesse-Biber. 2016. "Why so many 'rigorous' evaluations fail to identify unintended consequences of development programs. How mixed-methods can contribute". *Evaluation and Program Planning* 55 (2016): 155-162.

Table 1. Examples of key questions for evaluating gender and equity outputs and outcomes when applying each of the four types of evaluation

Evaluation approach	*Examples of key evaluation questions*
A. Policy evaluation	Assessing policies in terms of: • Relevance • Efficiency • Effectiveness • Impact • Gender responsiveness and equity • Sustainability To what extent can observed outcomes be attributed to the effects of the policy? How influential was donor agency advice in the formulation and implementation of national development strategies? For policies whose full results will not be seen for a number of years (until after the evaluation must be completed), what indicators can be used to estimate the likely success after a shorter period of time? Do policies contribute to gender equality and women's empowerment? Do policies contribute to the promotion of equality?
B. Formative evaluation	How likely is the programme design to achieve the different SDG development objectives? How effectively is the programme being implemented? Are any sectors of the target population being excluded or receiving less access to programme benefits? Are there any unintended outcomes (negative but also positive) that management must address? Do policies contribute to the promotion of gender equality and reducing inequalities?
C. Developmental evaluation	Are there mechanisms to ensure that all sectors of the target population are consulted? Are services and benefits reaching all sectors of the target population? Does the evaluation design identify and address all of the complexity dimensions of the project/programme? Does programme implementation have the flexibility to adapt to the changing context within which the programme operates? Do programmes contribute to gender equality and reducing inequality outcomes?

D. Summative evaluation	To what extent can a specific impact be attributed to the intervention?
	Did the intervention make a difference?
	How has the intervention made a difference?
	Will the intervention work elsewhere?
	What are the key factors (contextual, design, organization and coordination) that are important for successful replication?
	How simple or complex are the different dimensions of the programme on a complexity rating scale?[14]
	Is it necessary to use a complexity-responsive evaluation design?
	What are the main contextual factors affecting different programme outcomes?
	How are programme outcomes affected by problems of coordination among different stakeholders?
	Did the programme contribute to gender equality and to reducing inequalities outcomes?

The emergence of complexity theory

There is an increasing recognition that many development programmes are "complex" and that often this will require the utilization of "complexity-responsive" evaluation methodologies (see Chapter 3, section 3.4). There are some promising approaches being tested for the evaluation of complex programmes—such as theory of change-based approaches, contribution analysis, and outcome harvesting—but further work is needed. Consequently, the development of cost-effective and easy-to-apply complexity-responsive evaluation methodologies will be both a challenge and an opportunity for the evaluations of SDGs.

A greater focus on process and context

Many evaluations are designed to focus on outcomes and use methodologies that estimate change in intended outcomes over the life of the programme. This tends to divert attention for understanding what actually happens during programme implementation. This results in two serious limitations of the evaluation. First, if a programme fails to achieve some of its intended outcomes, it is often assumed that this is due to a weakness in the concept or design. However, it is often the case that at least part of the reason is due to problems during implementation. Very few programmes are implemented exactly as planned, so it is important to assess the relative importance of "design failure"

14 Bamberger, M., Vaessen, J., and E. Raimondo. 2016. *Dealing with Complexity in Development Evaluation: A Practical Approach*. Thousand Oaks, CA: Sage: Table 1.2.

and "implementation failure" in explaining why certain outcomes were not achieved. This requires the evaluation design to assess the effectiveness of the implementation process.

A second reason is that, in many cases, what happens during programme implementation is often more important than the achievement of specific outcomes. This is particularly true for programmes that promote gender equality and women's empowerment or that promote participation and inclusion.

The focus on complexity also stresses the need to analyse how programme implementation and outcomes are influenced by the economic, political, socio-cultural, ecological and other factors in the local, national and international context within which programmes operate. Many evaluation designs either ignore contextual factors or only introduce them in a non-systematic and anecdotal manner. Contextual factors are particularly important for EFGR evaluation because, in addition to the factors mentioned above, promoting women's empowerment or the inclusion of socially marginalized groups often threatens deeply held beliefs and practices, and there are often subtle but powerful forms of social control that must be identified and evaluated.

Continuous advance in gender and feminist methodology that can contribute to an equity-focused and gender-responsive approach

Since the launch of the MDGs, there have been a number of important developments in gender and feminist evaluation methodology that can contribute to the SDGs. While all of the approaches have roots going back earlier, they all reflect new emphases:

- Feminist theory stresses the importance of using a transformative lens to inform the evaluative process.[15] This prioritizes social justice and human rights as overarching ethical principles of an evaluation study. In addition to the assessment of how programmes do, or do not, transform power relations and women's greater freedom from mechanisms of social control, a transformative approach also stresses the importance of listening to multiple voices, all of which should have equal value in the conduct and interpretation of the evaluation. This is consistent with a human rights-based approach to evaluation, the rights-based underpinnings of the SDGs, and the goal of "No one left behind".

15 Mertens, D. M. and S. Hesse-Biber. 2013. "Mixed methods and credibility of evidence in evaluation". *New Directions for Evaluation*, 2013: 5–13. doi: 10.1002/ev.20053.

- The gender dimensions of social exclusion (see separate section below) is another new area of feminist research. This has also increased interest in the exploration of challenges and opportunities for women and men (and girls and boys) at different stages of the life cycle. The life cycle focus is potentially important for the analysis of the SDGs, as different interventions are targeted at different age groups and sectors of the population.

Intersectionality

In addition to evaluating progress on the different dimensions of gender and reducing inequalities, which is in itself quite challenging, the SDG framework stresses the complex interactions among the 17 SDGs. The dimensions of gender equality are influenced by other SDGs and, in turn, influence the achievement of them. For example, with respect to SDG-2 (end hunger and achieve food security), women's access to, and control over, productive resources and participation in decisions on family and community farming practices will have an important impact on a community's ability to increase agricultural output and increase food security. At the same time, when new agricultural technologies increase demand for women's inputs, this can have an important influence on women's empowerment. There are similar examples with respect to the position of excluded groups. Understanding these interactions requires the use of more sophisticated evaluation methodologies and application of intersectional analytical lens.

Using a social exclusion framework to assess "No one left behind"

Social exclusion is the process through which individuals or entire communities are systematically blocked from, or denied full access to various rights, opportunities and resources that are normally available to members of a different group and that are fundamental to social integration within that particular group.[16] The approach has been used in Europe for several decades, particularly to assess factors affecting the access of vulnerable groups to public services. Drawing on the extensive national databases available in most European countries the analysis examines the interaction among factors such as age, sex, ethnicity, education, and geographic location to determine the combination of factors affecting access to services for different groups at different stages of the life cycle.

16 Wikipedia article on social exclusion, July 2016.

In recent years, the social exclusion approach has been used by a number of UN agencies, including Food and Agriculture Organization, United Nations Children's Fund (UNICEF) and United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). Areas of research include: the gender dimensions of social exclusion, widows and orphans, the Roma population, food security and how social exclusions affects young children.

The social exclusion approach seems well suited for the assessment of the combination of factors that cause certain groups to be left behind. A number of social exclusion analysis frameworks have been developed, and the approach applies the concepts of intersectionality discussed in the previous section. One of the challenges in applying the social exclusion framework to EFGR evaluation of the SDGs is that many of the analytical approaches are very data intensive and are more difficult to apply in countries with more limited national statistics.

Advances in mixed method evaluation

Today mixed methods are widely accepted as one of the three main evaluation approaches (quantitative, qualitative and mixed method). Many would argue that all EFGR evaluations should adopt a mixed methods approach to combine an understanding of the lived experiences of women and men in different kinds of households, communities and economic activities (qualitative evaluation) with an estimate of the breadth and representativity of findings and the statistical significance of observed differences (quantitative evaluation).

It is important to recognize that mixed method evaluation is an integrated evaluation approach that requires the integration of quantitative and qualitative methods at all stages of the evaluation. It is much more than simply adding a few focus groups to a quantitative survey.[17]

The incorporation of values and different voices into the evaluation

Mixed method evaluations also recognize the importance of listening to multiples voices each with a different perspective on the programmes and processes being evaluated. This challenges the widely held belief that evaluation is an "objective" process that collects data

17 For references to mixed methods see: Teddlie, C. and A. Tashakkori. 2009. *Foundations of Mixed Methods Research.* Sage; Bamberger, M. 2016. "The importance of a mixed methods approach for evaluating complexity" in Bamberger, M., Vaessen, J., and E. Raimondo. 2016. *Dealing with Complexity in Development Evaluation: A Practical Approach.* Thousand Oaks, CA: Sage.

on a single reality that exists independently of the observer and of the context in which data is collected. Mixed methods and feminist evaluation recognize that different voices reflect different lived experiences and realities. This has a fundamental effect on how data is collected and interpreted and leads to questions such as "Whose voices count?" and "Whose reality will be studied?"

This perspective also recognizes that no evaluation can be "objective" or value-free. The decision of what to study, as well as who to listen to, implies value judgements that are not always made explicit. This presents important challenges for the design and presentation of the evaluation as some stakeholders who come from a quantitative background may question whether or not qualitatively oriented evaluations, based on capturing many voices, can be considered "professional" and "rigorous".

1.6 The implications of gender and reducing inequalities dimensions of the SDGs

Gender equality and reducing inequalities are both defined as cross-cutting themes for the SDGs. The evaluation of cross-cutting themes was not directly addressed in the MDGs and this adds considerably to the methodological difficulties of the evaluation design for the SDGs.

In the MDGs, gender equality was treated as a stand-alone goal, whereas for the SDGs, gender equality is both a cross-cutting theme and a stand-alone goal. The stand-alone goal (Goal 5) has nine targets that address many structural barriers to advancing women's rights. Gender-specific targets accompany these across many of the other goals.[18]

While gender equality is an important dimension of most SDGs, it is manifested differently for each goal. So while it is possible to develop a standard set of cross-cutting gender equality indicators, it may also be necessary to develop some specific indicators for different SDGs. Some of the specialized agencies evaluating areas such as energy, transport or climate change may have limited capacity or resources to conduct gender analysis and to develop these sector-specific gender indicators.

18 There are specific references to the need to target women or girls in SDG 1 (ending poverty), 2 (ending hunger), 3 (ensuring healthy lives), 4 (education), 5 (gender equality), 6 (water and sanitation, 8 (sustainable economic growth), 10 (inequality within and between nations), and 11 (cities and human settlements).

The evaluation of outcomes for reducing inequalities presents special challenges as the disaggregation of data to assess differential programme outcomes on different income groups or groups with different socio-cultural characteristics such as ethnicity, sexual orientation, disability, legal status or geographical location may challenge the data collection capacity of many agencies. There may also be political or cultural opposition to addressing some dimensions of social exclusion as in some contexts the special needs or even the existence of some of these groups may not be recognized by major political or social groups.[19]

1.7 Understanding the respective responsibilities and challenges for less developed, middle income and developed countries in the implementation of the evaluations

A central element of the follow-up and review processes (including evaluation) is that these will be voluntary and country-led, taking into account different national realities, capacities and levels of development, while respecting policy space and priorities.[20] Given different priorities, resources and evaluation expertise, it is likely that many countries may opt to focus on basic monitoring indicators or not to address gender equality and/or reducing inequalities in many sectors. Even where there is interest in assessing gender and reducing inequalities, the voluntary nature of the process means that countries may opt to use different approaches or will rely on existing data availability. Consequently, there are likely to be major challenges of comparability.

It is also possible that countries that do not address many of these issues will be those with more limited data availability. There is the risk of a selection bias with countries where gender equality and reducing inequalities may be particularly critical being those in which EFGR evaluations are not conducted or are more limited.

While these issues go beyond the scope of this publication, the definition of different regional and international collaborative mechanisms will be critical for strengthening the availability and quality of evaluation data from a broader range of countries and sectors. Similarly, UN treaty body system national reporting could also be systematically consulted when evaluating progress in the implementation of the SDGs. The

19 Bamberger, M. and M. Segone. 2011. "How to design and manage equity focused evaluations". New York, NY: UNICEF.

20 2030 Agenda, paragraph 74.

possibilities for South-South cooperation will play an important role, as will the contribution of the rapidly expanding voluntary professional evaluation associations.

1.8 Tailoring the evaluations of SDGs to national, sub-national and local contexts

The previous point has important implications for the tailoring of the evaluation efforts to national, regional and local contexts. A first set of issues concern the strengthening of national evaluation capacities in general, and specifically for EFGR evaluation. Traditionally, international support for strengthening evaluation capacity is coordinated through national planning and statistical agencies where the focus may be on the collection and analysis of basic quantitative data. Strengthening EFGR evaluation may involve different and more specialized agencies such as ministries of women's affairs or the gender units in line ministries. Much of the gender expertise is found in civil society organizations and women's advocacy groups, some of which do not work closely with central government agencies, so an effort may be required to strengthen the collaborative mechanisms.

A challenge will be to strengthen the capacity of existing M&E systems to address SDGs. In addition to training and other kinds of technical support, this will involve finding ways to strengthen cooperation between government and civil society evaluation departments. In addition to broadening the range of evaluation resources and technical expertise, many of the civil society organizations conduct evaluations at the community and local level, which is very important for understanding the social mechanisms that maintain gender inequalities and other forms of social exclusion.

The UN Development Group has prepared a useful guidance note on tailoring the 2030 Agenda to the national context, which also provides some general guidance on how to develop country M&E systems.[21] The guidance proposes a four-step process:

21 UN Development Group. 2015. "Guidance note: On supporting the tailoring of the 2030 Agenda for Sustainable Development to nation contexts". Version 1.0, September 14, 2015.

1. Reviewing existing strategies and plans and identifying required areas of change;
2. Making initial recommendations to the leadership of the national government;
3. Setting nationally relevant targets;
4. Formulating strategies and plans using systems thinking.

Section B7 of the UN Development Group Guidance provides guidelines for developing monitoring, reporting and accountability covering four topics:

1. ***Indicator development and data collection.*** It is recommended that countries should follow the progress of the expert group on SDG indicators[22] and adapt these to the national context.
2. ***Disaggregating data to ensure that "No one is left behind".*** It is important to work with national statistical offices to ensure that all sources of data relating to household economic and social indicators are disaggregated by sex and other relevant categories. This will make it possible to compare information on economic indicators and access to services on vulnerable groups with the averages for the total population. SDG sub-goal (17.18) is devoted to data disaggregation.
3. ***Developing monitoring and reporting systems that cover all relevant SDG sub-goals.***
4. ***Putting in place review processes and mechanisms for collection and analysis of M&E data.***

22 See for example, the report on open consultations on green indicators (4-7 November 2015), which developed a preliminary list of indicators for each of the sub-goals for each SDG. Of particular relevance for the present purpose are the indicators proposed for SDG 5 and 10. http://unstats.un.org/sdgs/iaeg-sdgs/open-consultation-2

2. SDG follow-up and review mechanisms

2.1 The SDG framework for follow-up and review

The 2030 Agenda made a commitment to ensure a systematic follow-up and review of the SDGs that would be "robust, voluntary, effective, participatory, transparent and integrated" and that would "make a vital contribution to implementation and will help countries to maximize and track progress in implementing the 2030 Agenda *in order to ensure that no-one is left behind*".[23] Box 2 summarizes the basic principles on which the approach is based.

2.2 The role of different actors at the national and subnational levels

National governments are the key agencies responsible for the implementation of the evaluations of SDGs within each country. As the reporting systems and evaluations are voluntary, the commitment of governments will be critical, particularly as they have to decide how to prioritize their limited financial and technical resources among many different development priorities—all supported by different groups of international and national stakeholders. Given the broad scope of the SDGs, almost all government agencies will potentially be involved and the national government will play an important coordinating role. One of the challenges will be to avoid the "silo mentality" (seen in many of the MDG M&E activities) whereby each sector agency works on its own sector-specific studies with very little coordination between sectors.

The evaluations will also have important local dimensions and the national government must also coordinate with agencies at these levels.

At the national level, donor agencies, UN agencies, civil society, advocacy groups and foundations can all play important roles in determining the research/evaluation agenda. There is always a danger that each donor agency, civil society and UN agency will conduct their own studies, often with only limited coordination, significant duplication and only limited comparability of data. The Inter-Agency Expert Group on SDGs (IAEG-SDGs) is seeking to avoid these issues through

23 2030 Agenda, paragraph 74.

BOX 2. The principles of the SDGs follow-up and review mechanisms

SDGs follow up and review mechanisms at all levels will be guided by the following principles:

a. They will be voluntary and country-led, will take into account different national realities, capacities and levels of development, and will respect policy space and priorities. As national ownership is key to achieving sustainable development, the outcome from national level processes will be the foundation for reviews at regional and global levels, given that the global review will be primarily based on national official data sources.

b. They will track progress in implementing the universal goals and targets, including the means of implementation, in all countries in a manner that respects their universal, integrated and interrelated nature and the three dimensions of sustainable development.

c. They will maintain a longer-term orientation, identify achievements, challenges, gaps and critical success factors and support countries in making informed policy choices. They will help mobilize the necessary means of implementation and partnerships, support the identification of solutions and best practices, and promote coordination and effectiveness of the international development system.

d. They will **be open, inclusive, participatory and transparent for all people** and will support the reporting by all relevant stakeholders.

e. They will **be people-centered, gender-sensitive, respect human rights and have a particular focus on the poorest, most vulnerable and those furthest behind.**

f. They will build on existing platforms and processes, where these exist, avoid duplication and respond to national circumstances, capacities, needs and priorities. They will evolve over time, taking into account emerging issues and the development of new methodologies, and will minimize the reporting burden on national administrations.

g. They will **be rigorous and based on evidence, informed by country-led evaluations and data that is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts.**

h. They will **require enhanced capacity-building support for developing countries, including the strengthening of national data systems and evaluation programmes**, particularly in African countries, least developed countries, small island developing states, landlocked developing countries and middle-income countries.

i. They will benefit from the active support of the UN system and other multilateral institutions.

Source: The 2030 Agenda for Sustainable Development paragraph 74. Bold text by author.

developing recommended standard indicators, but national governments will have a major role in ensuring these guidelines are followed.

Civil society organizations will play an important role in the evaluations at both the national and local levels and their contribution will be critical in ensuring a truly inclusive consultation and participatory approach. While many governments collect data on local communities and are willing to involve these communities in the data collection process, government agencies are often less willing to involve them in the interpretation of the findings and in the discussion of the policy implications of the findings. Civil society, and particularly women's rights and feminist groups will have an important role to play in ensuring that the voices of local communities and marginalized groups are heard.

2.3 Strengthening and using existing M&E systems at the local, national and international levels

The IAEG-SDGs has developed a set of basic indicators for all of the SDGs and sub-goals, which it is hoped all countries will follow. The challenge will be to assist countries in the collection and analysis of these indicators, particularly given the significant differences in the capacity (financial and technical) of different countries to collect the data. In addition, multiple development priorities mean that countries will vary in terms of their incentives to commit scarce resources to the evaluations of SDGs.

In 2015, the UN Development Programme conducted a review of national evaluation capacities in 43 countries.[24] This showed there are major differences in budgets, capacities and how evaluations are conducted and used. There is no single, uniform structure that all countries use, and control of the selection, implementation and use of evaluations is located in different agencies in different countries. There are also considerable variations in the extent to which there is a national evaluation policy (NEP) and a centralized evaluation system.

A 2015 UN Women study found that countries also varied in terms of the kinds of data they collected on gender and ethnic and cultural issues.[25] In many, but not all countries, data is disaggregated by sex, but often gender analysis does not go beyond that. In only a few

24 UNDP. 2015. "Towards a baseline study: Insights on national evaluation capacities in 43 countries". Independent Evaluation Office. New York, NY: UNDP.

25 UN Women. 2015. "Monitoring gender equality and the empowerment of women and girls in the 2030 Agenda for Sustainable Development: Opportunities and challenges". New York, NY: UN Women.

countries is data regularly collected on ethnic and cultural issues, and it is to be presumed that even less data is regularly collected on vulnerable groups. There are also instances where household rosters are compiled covering all household members but where only aggregated data (not broken down by sex) is published at the household level. It may be possible to reanalyse existing survey data to obtain sex disaggregated data for at least some sectors. There are also sectoral differences so that school data is always disaggregated by sex, but it is common to find that data on road accidents is not. There are also major differences in the quality and completeness of sex disaggregated data or of data affecting women. For example, police records on violence against women tend to be incomplete and inaccurate—if they are collected at all.

It is common to find that M&E systems are stronger in sectors that are government priorities or that receive strong donor support, but frequently there is no integrated system covering all development sectors. A number of challenges face efforts to strengthen national evaluation capacity in areas such as EFGR evaluation:

- Many countries have weak evaluation structures, which could not support the additional burden of collecting gender-related data
- Lack of financial and technical resources
- Difficulties and costs of data collection on many gender-related issues
- Reluctance of many agencies to conduct gender-responsive evaluations for the following reasons:
  - Many agencies do not see gender as relevant or useful. Many programmes (such as transport, energy, water, banking and finance and trade) are believed to be "gender neutral".
  - The methodologies are unfamiliar to many agencies and are perceived as difficult to use.
- Even if evaluation units are supportive, collecting gender-related data will often be an additional burden on operational units, and evaluation units are often reluctant to ask for more data (and time) from their over-burdened operations colleagues

Annex 3 presents cases illustrating ways that multiple stakeholder partnerships have helped strengthen national M&E systems.

2.4 Steps towards strengthening the capacity of existing M&E systems to conduct gender analysis and to address inequality issues

The following are approaches that may be applicable in different contexts:

1) Draw on existing M&E guidance documents focusing on gender, human rights and reducing inequalities such as the UN Evaluation Group "Integrating human rights and gender equality in evaluation"[26] and the UNICEF equity focused evaluation approach.[27]
2) Carefully review existing M&E systems at the national and sector levels to determine:
    a. The kinds of data on gender and reducing inequalities that are-already analysed and published.
    b. The kinds of data that have been collected but not analysed or published. Discuss with M&E staff the feasibility of conducting additional analysis to generate disaggregated data. Be sure to check what would be the cost, resources or political issues in conducting this analysis and what kinds of support would be necessary to make this happen.
    c. Discuss with national and sector statistical offices and M&E staff the feasibility of the following options (which may vary by sector and region):
        i. The possibility of collecting disaggregated data by sex or indicators of inequality in future evaluations.
        ii. The possibility of including a few additional questions on gender or inequality: check what kinds of data agencies might be willing to collect (e.g., time use) and which would be too difficult or sensitive to collect (e.g., violence against women, information on ethnic minorities or vulnerable groups).
    d. Willingness to expand sample size if additional funding could be mobilized.
    e. Willingness to allow teams working on gender and reducing inequalities to administer a special module (e.g., on time use or women's opinions on projects) to a sub-sample of respondents.
    f. Cooperation on a joint evaluation combining a quantitative survey with in-depth case studies to be administered by teams on a sub-sample. This would have two objectives: to increase

26 UN Evaluation Group. 2013. "Integrating human rights and gender equality in evaluation: Guidance document".

27 Bamberger and Segone 2011.

the statistical representativity of the in-depth samples and to gain wider support for the gender analysis by building these kinds of alliance.

3) Explore ways to strengthen cooperation between government and civil society organizations (with expertise in gender and reducing inequalities research) in the design and implementation of evaluations. This may require a significant resource and time commitment as well as careful negotiation in countries or sectors where relationships are sensitive. Often country reporting on UN treaty body implementation offer good analysis of the human rights and gender situation as seen by civil society actors (as through shadow reports).
4) Evaluation capacity development initiative—There are many modalities including: sending staff to conferences, and short or longer training programmes; bringing in consultants to work with staff; and alliances with agencies that have this expertise or with national and regional evaluation associations. Some of the skill development areas on which the evaluation capacity development programmes should focus include:
    a. EFGR evaluation frameworks so that evaluators understand what kinds of data should be collected and why;
    b. Methodologies for the collection and analysis of gender and inequality focused data;
    c. Strategies for identifying the availability of sex-disaggregated and equality-focused data, how to assess quality and practical ways to put in place a disaggregation strategy;
    d. How to market an EFGR evaluation strategy;
    e. Basic gender-responsive evaluation designs and how they can be put in practice;
    f. How to mainstream EFGR evaluation in existing evaluation policies and national M&E frameworks.

2.5 Assessing the quality and coverage of data on gender equality and data on reducing inequalities in the SDGs

SDG coverage of gender equality and women's empowerment

From a gender perspective, the SDGs, particularly SDG-5 but also the gender-related targets in the other goals, represent a significant step forward from the MDGs, covering for the first time core areas of women's rights and women's empowerment. There is a strong realization this time that gender equality and the empowerment of women and girls is essential to achieve progress across all the goals and targets.

The historic and unprecedented ambition set out in the 2030 Agenda must be matched by an equally ambitious drive to ensure its implementation. This will require an adequately resourced and authoritative monitoring and accountability framework. Robust indicators and quality data are of critical importance and will to a large extent determine whether or not policy efforts are marshalled and the goals and targets are achieved or missed. Building integrated information systems that address gender equality in all its dimensions will provide a credible evidence base that can inform such policies and catalyse actions.

SDG-5 speaks specifically to governments' commitments to: end all forms of discrimination against women and girls; eliminate all forms of violence against women and girls; eliminate all harmful practices, such as child marriage and female genital mutilation; recognize and value unpaid care and domestic work; ensure women's full and effective participation and leadership at all levels of decision-making; ensure universal access to sexual and reproductive health and reproductive rights; provide women's equal rights to economic resources; promote women's empowerment through the use of technology; and strengthen policies and laws for the promotion of gender equality.

The full text of SDG-5 (Achieve gender equality and empower all women and girls) is presented in Box 3.

The IAEG-SDGs has identified 230 indicators for global monitoring, approximately one quarter of which are gender-related. The priority is ensuring that countries increase their capacity to collect data on the gender indicators in the official IAEG-SDG list. The challenges of effectively monitoring the SDGs from a gender perspective cannot be overstated. Of the 14 proposed indicators to monitor SDG-5, for example, there are only three (referred to as Tier I indicators) for which internationally accepted standards for measurement exist and for which data

BOX 3. SDG-5: Achieve gender equality and empower all women and girls

5.1 End all forms of discrimination against all women and girls everywhere

5.2 Eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking and sexual and other types of exploitation

5.3 Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation

5.4 Recognize and value unpaid care and domestic work through the provision of public services, infrastructure and social protection policies and the promotion of shared responsibility within the household and the family as nationally appropriate

5.5 Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life

5.6 Ensure universal access to sexual and reproductive health and reproductive rights as agreed in accordance with the Programme of Action of the International Conference on Population and Development and the Beijing Platform for Action and the outcome documents of their review conferences

5.7 Undertake reforms to give women equal rights to economic resources, as well as access to ownership and control over land and other forms of property, financial services, inheritance and natural resources, in accordance with national laws

5.8 Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women

5.9 Adopt and strengthen sound policies and enforceable legislation for the promotion of gender equality and the empowerment of all women and girls at all levels

are regularly collected by most countries. Of the remaining 11 indicators, some have internationally accepted standards but data collection by most countries is largely irregular and for others significant methodological work is needed as international standards do not yet exist and most countries do not regularly collect the data needed.

The global profile of gender statistics and the recognition of the need to fill critical gaps in data coverage have increased exponentially. The requirements of monitoring the SDGs from a gender perspective demand a significant scaling-up of this work in order to realize the ambition set out in the SDGs. In this context, UN Women is taking a leading role in addressing gaps in gender statistics, working with other UN agencies as well as national statistical offices and civil society organizations.

SDG inclusion of reduction of inequality indicators at the country level

In addition to SDG-10 there are 10 other SDGs that include at least one indicator of inequality. For this discussion, we consider an indicator as referring to reduction of inequality when it specifically refers to vulnerable or disadvantaged groups (including a reference to the bottom proportion of the income distribution (for example, the bottom 40 per cent). A reference to "all people" (e.g., "ensure that all boys and girls complete free primary and secondary education" SDG-4.1) is not considered an indicator of inequality reduction. The reason is that quantitative increases in access to services will often not reduce the gap between the most vulnerable groups and the rest of the population. Consequently, it is important to have goals that specifically focus on the vulnerable groups and the inequality gap.

Some indicators do specifically focus on the poor and vulnerable (e.g., SDG-1.4 "Ensure that all men and women, in particular the poor and the vulnerable have equal access to economic resources …"). While it could be argued that the analysis of all indicators could easily use techniques such as quintile analysis to compare access to benefits for each indicator, it is important to ensure that the equality focus is specifically stated in the indicator as otherwise it is likely that in many cases the analysis *will not specifically address the inequality dimension*.

BOX 4.SDG-10: Reduce inequality within and between countries

10.1 By 2030, progressively achieve and sustain income growth of the bottom 40 per cent of the population at a rate higher than the national average

10.2 By 2030, empower and promote the social, economic and political inclusion of all, irrespective of age, sex, disability, race, ethnicity, origin, religion or economic or other status

10.3 Ensure equal opportunity and reduce inequalities of outcome, including by eliminating discriminatory laws, policies and practices and promoting appropriate legislation, policies and action in this regard

10.4 Adopt policies, especially fiscal, wage and social protection policies, and progressively achieve greater equality

10.5 Improve the regulation and monitoring of global financial markets and institutions and strengthen the implementation of such regulations

10.6 Ensure enhanced representation and voice for developing countries in decision-making in global international economic and financial institutions in order to deliver more effective, credible, accountable and legitimate institutions

10.7 Facilitate orderly, safe, regular and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies

10.8 Implement the principle of special and differential treatment for developing countries, in particular least developed countries, in accordance with World Trade Organization agreements

10.9 Encourage official development assistance and financial flows, including foreign direct investment, to states where the need is greatest, in particular least developed countries, African countries, small island developing states and landlocked developing countries, in accordance with their national plans and programmes

10.10 By 2030, reduce to less than 3 per cent the transaction costs of migrant remittances and eliminate remittance corridors with costs higher than 5 per cent

3. Relevant principles of the 2030 Agenda for a "No one left behind" evaluation

3.1 Framing the evaluation within the SDG principles

It is important to ensure that the evaluation strategies and approaches are fully consistent with the SDG principles for follow-up and review. This will ensure that the dimensions of gender and reducing inequalities are incorporated as an integral component of all of the SDG evaluations, and are not considered as special, stand-alone topics that are only of interest to gender specialists. At the time of writing, the SDGs has not yet developed evaluation guidelines, and the current principles relate to the follow-up and review framework described in Chapter 2 of this guidance.

3.2 Gender equality, reducing inequalities and ensuring "No one left behind"

Gender equality, reducing inequalities and ensuring "No one left behind" are considered as distinct but linked core principles of the SDGs. Evaluations of SDGs policies should incorporate these principles throughout the evaluation in order to help address multiple causes of discrimination and exclusion.

As a result of the work of gender equality advocates, gender equality is reflected throughout the 2030 Agenda, including in the declaration; goals, targets and indicators; means of implementation; global partnership; and follow-up and review. Gender equality and women's empowerment is recognized as a stand-alone goal (Goal 5) and also as a cross-cutting issue that is mainstreamed throughout the SDG.

Gender equality in the 2030 Agenda is also clearly mentioned in the Preamble where the SDGs "seek to realize the human rights of all, and to achieve gender equality and the empowerment of all women and girls. They are integrated and indivisible…". There is a call for integrating gender equality into SDG policies as its key contribution to progress across all goals and targets. The implication for evaluation is to strengthen gender-responsive approaches that will contribute to deeper analysis of social norms and behaviour than is possible with

conventional quantitative analysis and sex-disaggregation of data. The goal of gender-responsive evaluation is to:

1. Assess the degree to which gender and power relationships—including structural and other causes that give rise to inequities, discrimination and unfair power relations—change as a result of an intervention using a process that is inclusive, participatory and respectful of all stakeholders (rights holders and duty bearers)
2. Provide information on the way in which development programmes are affecting women and men differently and contributing towards achievement of these commitments
3. Help promote social change by using the knowledge produced from an evaluation for better development programming that promotes gender equality, women's empowerment and human rights in a sustainable manner

While understanding the dynamics of gender equality applies many of the same principles used in the analysis of economic or geographic inequality or exclusion (i.e., quintile analysis, public expenditure incidence analysis and social exclusion analysis), there are additional and more complex mechanisms explaining why women (or men) may be excluded from access to public resources, labor markets or participation in political decision-making. Many of these relate to the range of legal, institutional, economic, political, social and psychological factors, which together constitute the complex web of social control of the behaviour of women and men in a given society. The implication is that while the first phase of gender equality analysis can be based on the conventional tools of economic equality analysis, many studies must go beyond these techniques to dig deeper into how society controls the behaviour and opportunities for women and men.

One important issue for data collection and analysis is that much equality analysis takes the household as the unit of analysis, assuming that all household members have equal access to resources. A basic principle of gender analysis is that, in most societies, there is an unequal distribution of food, productive resources and access to decision-making. Consequently, gender equality analysis will frequently require the development and use of special data collection tools that permit this disaggregated data collection and analysis.

The SDG principle of "No one left behind" is based on a concept of *equality*. This recognizes the need to go beyond aggregate indicators, which only estimate the proportion of the population who have benefited from a particular intervention, such as: the proportion of girls

attending secondary school, or the proportion of households with access to water, electricity or primary health services. There is accumulating evidence that aggregate indicators of progress can conceal the fact that some marginal or vulnerable groups are left behind. Sadly, a rising tide does not mean that all boats are lifted up. The goal of the SDGs in reducing inequalities is to:

- Identify groups who have been left behind
- Understand why this has happened
- Identify strategies to promote more inclusive approaches that include these groups

3.3 Sustainable development: Sustainability and resilience

Among SDG documents consulted, there does not appear to be a definition or a conceptual framework explaining the concept of sustainability or the process through which sustainability is to be achieved or progress to be assessed. While it would be very difficult to develop a sustainability-responsive theory of change for achieving the SDGs, it might be possible to do this for individual SDGs (or even for particular sub-goals). Box 5 gives a short definition of sustainable development (taken from the Bruntland 1987 report) as "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs".[28] This must incorporate at least four dimensions: ecological, economic, political and socio-cultural sustainability.

While the SDGs only focus directly on "sustainability", a complementary concept that should also be addressed is "resilience". Resilience refers to "the ability of a system, entity, community or person to withstand shocks while still maintaining its essential functions and to recover quickly and effectively from catastrophe." While sustainability focuses on the ability of a system to maintain equilibrium, resilience focuses on the ability of a system to learn from shocks and stresses and to learn to adapt. In some ways, resilience is a more dynamic concept as it recognizes that environments are constantly changing. The concept of resilience has been used quite widely to understand how community organizations, including women's organizations, have learned to cope with stress.[29]

28 Taken from the Bruntland 1987 report.

29 Zolli, A. 2012. *Resilience: Why Things Bounce Back*. New York, NY: Free Press.

Evaluating sustainability and resilience

Evaluating sustainability and resilience requires a very different methodological approach compared to the conventional evaluations of programme outputs and outcomes. Both sustainability and resilience involve assessing the ability of communities or other entities to respond to and learn from shocks, stresses and processes of change, which may occur over long periods of time and are usually unpredictable. Also, it is not possible to assess how resilient and sustainable an entity is until some time after a dramatic shock or other disturbing event has occurred. In some cases, it may be necessary to wait

BOX 5. Defining sustainability and resilience

Sustainability

According to the World Commission on the Environment and Development, sustainable development is "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs."

Sustainability was first introduced by ecologists. Herman Daly identified three dimensions of ecological sustainability:

1) Renewable resources: the rate of harvest should not exceed the rate of regeneration

2) Pollution: the rates of waste generation from projects should not exceed the assimilative capacity of the environment

3) Nonrenewable resources: the repletion of the non-renewable resources should require comparable development of renewable substitutes for that resource

When the concept was broadened to define sustainable development, a number of additional interconnected domains have been proposed including:

- Ecological sustainability
- Economic sustainability
- Political sustainability
- Social or cultural sustainability

Resilience

"The ability of a system, entity, community or person to withstand shocks while still maintaining its essential functions and to recover quickly and effectively from catastrophe".

"Resilience is what enables people to survive, and thrive."

Sources:World Commission on Environment and Development 1987; Daly 1990; and Bamberger and Kumar 2013.

until the next stress or shock to assess whether or not the ability to respond has improved.Recognizing the long time periods involved in the assessment is important, because many evaluations have an untested belief that, for example, strengthening a certain kind of community organization will increase resilience without having any supporting evidence. Consequently, assessments sometimes claim that resilience or sustainability have increased when this may not, in fact, be the case. The evaluation methodologies for assessing sustainability and resilience are discussed in Chapter 4.

The importance of sustainability and resilience for equity-focused and gender-responsive evaluations

Many programmes to address gender equality face challenges from social, political and economic forces that resist changes to deeply entrenched privileges, values and beliefs. For this reason, gender analysts refer to "one step forward and two steps back" when initial progress is installed by these kinds of resistance. Successful programmes must have the resilience to learn from shocks and setbacks and often learn new approaches or more effective ways to overcome resistance.

3.4 Interlinkages across SDGs: Complexity and complex development programmes

As development programmes, including many programmes to implement SDGs, increase in size, the scope of their goals, and the number of stakeholders involved, they become more complex. Larger programmes also become increasingly affected by political, economic, ecological, socio-cultural and other elements of the local, regional, national and international contexts within which they operate. The interactions among all of these factors, make it increasingly difficult to track linkages between the multiple inputs and the equally numerous outcomes (some intended and many others unanticipated or often not even desired). In addition to the problems of measuring the multiple outcomes, it is extremely difficult to identify causal relations and to assess what contribution the programme (or often multiple programmes) have made to these outcomes.

For all of these reasons, evaluators and managers are coming to recognize that most conventional evaluation designs have serious limitations on their ability to evaluate complex programmes. At this point in time, there are no well-established methodologies for the evaluation of complex programmes, but a number of promising approaches are

beginning to emerge—although none are yet widely used or tested.[30] A challenge for SDG evaluators will be to draw on the experience of the network of research and evaluation agencies involved in the SDGs to identify, test and operationalize approaches that can address the challenges of complexity, while at the same time being sufficiently simple and economical to be applicable throughout the programmes implementing the SDGs.

The dimensions of complexity

It is useful to think of complexity as having four dimensions, all of which interact with each other. These are summarized below:

Dimension 1: The nature of the intervention

As the goals and scope of a programme become broader and more ambitious, the nature of the programme or intervention tends to become more complex. Some of the sub-dimensions that determine the level of complexity include:

- The size of the programme and of the target population, including the most vulnerable
- The number of components or services provided
- The technical and social complexity of the programme[31]
- The number of programme objectives and the clarity of their definition
- Is the programme design well tested or relatively new and still experimental?
- The duration of the programme (the longer programmes last, the more complex they become as they are more likely to be affected by administrative or political changes or by the evolving context in which they operate)
- The need to target particular, difficult-to-reach groups, such as the "bottom 40 per cent" that SDGs seek to include

30 USAID's *gender aware* evaluation (USAID 2013, Patton's Developmental Evaluation 2011) presents a range of strategies for dealing with complexity (see, for example, the 10 design examples starting on page 315); Bamberger et al. 2016, propose an "unpacking strategy" for breaking complex programmes into a set of easier to evaluate component (which must then be reassembled to understand how effectively the programme addresses broader contextual factors); and Funnell and Rogers 2011, discuss how logic models and theories of change can address complexity.

31 Technical complexity refers to development programmes where the design requires specialized technical knowledge in fields such as medicine, communication technology or hydraulic engineering. In contrast, social complexity refers to programmes that require introducing behavioural change or promoting cooperation among groups that either have not worked together before or where there has been a history of conflict.

Dimension 2: Stakeholders and the institutional framework

As programmes become larger and more ambitious, the number and diversity of stakeholders (i.e., including the most vulnerable) and implementation agencies tends to increase, as well as the range of interactions among them. Some of the sub-dimensions that determine the level of complexity include:

- The number of international, national and local stakeholders
- The number of agencies involved in programme implementation
- The number of agencies or consulting groups involved in the M&E of different components
- The number of agencies providing funding and the clarity of definition and coordination of funding arrangements
- The effectiveness of communication and coordination among funding agencies
- Complexity is further increased when a wider range of vulnerable and difficult-to-reach groups are involved, including women and women's organizations who are often voiceless

Dimension 3: The range of contextual factors and their level of influence on programme implementation and outcomes

Programmes operate in local, regional, national and international contexts where they may be influenced by economic, political, institutional, socio-cultural, historical, demographic and ecological factors. Programmes also vary in terms of how much they are influenced by these contextual factors (the level of contextual dependence). Some of the sub-dimensions include:

- How well are contextual factors and their level of influence understood
- How much influence does each contextual factor have
- The level of contextual dependency

Dimension 4: The nature of causality

In small projects with a low level of programme complexity, relatively simple institutional arrangements and a low level of contextual dependence, it is possible to trace and evaluate a direct causal relationship between a programme intervention (e.g., drinking water, scholarships for girls to attend secondary school) and the intended outcome (e.g., lower rates of diarrhea, higher rates of girl's enrolment). As programmes become more complex in terms of the three previous dimensions, the number of inputs increases (often operating differently in

different communities or regions), the number of intended and unintended outcomes also increases, and the influence of different stakeholders and institutional arrangements becomes more complicated, as well as the number of contextual factors. Consequently, it becomes increasingly difficult, or in many cases impossible, to determine direct causal relationships. It is quite common to find that the same package on programme inputs will produce different outcomes in different communities or regions, and that a given outcome can be produced by different combinations of inputs. As the level of complexity increases, the causal linkages are no longer linear and processes become recursive and non-proportional.

A practical and widely used approach in these situations is contribution analysis[32] and variations such as outcome harvesting[33] and outcome mapping.[34] These approaches develop a theory of change to develop and test the most credible storyline. This is tested with all of the available evidence. Some of these approaches are developed at the start of the programme while others are used either prospectively or retrospectively. Approaches also differ in terms of whether or not they seek to describe and test the intended project design (contribution analysis) or whether or not they have a broader focus that seeks to harvest all of the opinions of stakeholders on all of the changes that occurred, including those that were not intended, or even desired, by the project design (outcome harvesting). All of these approaches recognize that in complex programme contexts, it is not possible to use attribution analysis, and the best that can be expected is to assess the contribution of different interventions to the observed changes.

A familiar situation in gender analysis is "one step forward and two steps back" where, for example, an intervention to increase women's access to financial resources and technical advice may produce some short-term improvements in women's economic position, but there may then be a negative reaction if powerful groups believe that "women have gone too far". So tracing the influence of, for example, a microcredit programme may require following the programme over a number of years and tracking the complex trajectory with all of the

32 Mayne, J. 2011. "Contribution analysis: Addressing cause and effect". In Forss, K., Marra, M., and R. Schwartz. Eds., *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers.

33 Wilson-Grau, R., and H. Britt. 2012. "Outcome harvesting". Cairo, Egypt: The Ford Foundation's Middle East and North Africa Office. Available at: http://www.outcomemapping.ca/resource/resource.php?id=374.

34 Earle, S., Carden, F. and T. Smytlo. 2001. "Outcome mapping: Building learning and reflection into development programmes". Ottawa: International Development Research Center.

advances and setbacks. Gender outcomes are also very culturally sensitive so that outcomes of the same programme may be relatively successful in one village but much less so in another neighboring community. The following are some of the sub-dimensions:

- Are there single causal pathways linking inputs and outcomes, or are there many different pathways linking multiple inputs and outcomes?
- The nature of causal linkages[35]: Are they linear or non-linear? Is change proportional or non-proportional? Are there recursive patterns?
- The number of institutional and contextual variables influencing the relations between inputs and outcomes

A fifth dimension: The complexity of the evaluation itself

While the previous four dimensions refer to the complexity of the programme and the context within which it operates, it is sometimes useful to also assess the complexity challenges involved in the design and conduct of the evaluation itself. While the complexity of the evaluation tends to increase in line with the complexity of the programme, this is not always the case. Often the evaluation may only address one aspect of the programme (for example, only examining the relationship between scholarships and girls enrolment while ignoring contextual factors and institutional complexities of girls' school enrolment and performance). In other cases, the evaluation of a (seemingly) relatively simple programme may become more complex either because of difficulties in obtaining data or because the evaluation digs deeper into processes of behavioural change. Some of the factors making the evaluation process complex include:

- Lack of standard and consistent M&E systems with comparable data
- Barriers (political, methodological or logistical) to the collection of data
- The nature of causality and how this affects the complexity of the evaluation design
- Does the evaluation design use the more challenging mixed method designs?

35 Linear causal pathways refer to situations where positive or negative changes in inputs produce similar changes in outputs. Non-proportional change is a form of non-linear change but is used in situations where either a small change in an input suddenly begins to produce very large changes in outcomes (referred to as a "tipping point"), or where very large changes in inputs produce small or no change in outcomes (referred to as "inertia").

Table 2 presents a checklist identifying some of the sub-dimensions that together determine the level of complexity of each of the four dimensions (from low = 1 to high = 5 on a 5-point scale).

Table 2. Checklist for assessing the level of complexity of a gender-responsive programme

	Complexity rating						
Dimensions	***Low***		***1***	***2***	***3***	***4***	***5***	***High***
Dimension 1: The nature of the intervention								
1.1 Objectives	Few and relatively clearly defined							Multiple, broad and often not clearly defined
1.2 Size	Affecting small population							Affecting large population
1.3 Stability of programme design	Relatively stable							Emergent design
1.4 Is the programme design well tested and clearly defined?	Well tested and used many times							Relatively new and untested
1.5 Services or components	Relatively few							Large number
1.6 Technical complexity	Low							High
1.7 Social complexity	Low							High
1.8 Duration	Clear start and end date							No clear end date and sometimes no clear start date
1.9 Does the programme challenge established systems enforcing unequal gender relations?	Does not challenge any established systems							Programme designed to change legal, political or cultural systems
Dimension 2: Institutions and stakeholders								
2.1. Budget	The use of the funds is clearly defined							General budget support with no clear definition of services to be funded
2.2 Funding and implementing agencies	Relatively few							Large number
2.3 Stakeholders	Relatively few and with similar interests							Many and diverse
2.4 Does the programme give voice to new stakeholders representing women and disadvantaged groups?	The programme only works with established stakeholders							The programme seeks to promote new stakeholder groups

Dimensions	*Complexity rating* *Low* *1 2 3 4 5* *High*						
Dimension 3: The context within which the programme operates							
3.1 Context dependence	Relatively independent of the context (robust)						Programmes are strongly influenced by contextual factors
3.2 Strength of contextual factors	There are few powerful contextual factors						Many strong contextual factors
3.3 Is the programme likely to be influenced by norms and power dynamics affecting the status of women?	The programme is not likely to be influenced by norms and power dynamics						The nature of the programme makes it likely it will be influenced by norms and power dynamics
Dimension 4: Causality and change							
4.1 Causal pathways	Single and linear causal pathway						Multiple causal pathways (non-linear, interconnected, recursive feedback loops)
4.2 Agreement on appropriate actions to address problems	Relatively high agreement						Relatively low agreement
4.3 Certainty on outcomes	Relatively high degree of certainty						Low degree of certainty
4.4 Complex processes of gender-related behavioural change	The programme is not likely to involve complex processes of behavioural change						Complex processes of behavioural change are likely to occur

Source: Adapted from Bamberger et al. 2016, Chapter 1.

The importance of a complexity framework for understanding the SDGs

This section illustrates how complexity is likely to apply to SDGs, especially to those programmes addressing gender and reducing inequalities. It is argued that the requirement of gender-responsive SDGs to challenge a wide range of legal, organizational, political, economic and cultural barriers, the complex environments in which many programmes operate, and the complex processes of behavioural change they seek to promote, means that gender-focused programmes are likely to rate high on many of the indicators in this checklist.

Dimension 1: The nature of the national programmes addressing SDGs

Given the broad range of the SDG goals and the fact that multiple programmes are being implemented in each country by hundreds or sometimes thousands of different agencies, the overall SDGs in each country are likely to be rated high on overall complexity. There are, of course, many small and simple programmes each providing one or a few clearly defined services in a small number of communities, but when all programmes are combined, the level of complexity rapidly increases. Referring to the checklist in Table 2:

1.1 Objectives: The SDGs have multiple objectives. There are 169 targets, and many countries will identify country-specific targets. While some targets are clearly defined, many others are not, or there is a lack of consensus among stakeholders.

1.2 Size: While some programmes are quite small and localized, many cover whole regions or the whole country and affect very large populations.

1.3 Stability of programme design: The SDGs combine many well-tested and stable programmes with many new and innovative programmes. SDG-5 is likely to contain a large proportion of new and experimental programmes promoting gender equality and women's empowerment.

1.4 Is the programme design well tested and clearly defined?: The SDGs combine programmes with clearly defined designs and other experimental or dynamic and changing designs.

1.5 Services and components: While some programmes have few components, others have multiple components and services. Given the complexity of promoting changes in deeply entrenched practices and beliefs, many gender equality programmes involve coalitions of organizations each providing different but complementary programmes.

1.6 Social complexity: In contrast, the level of social complexity of gender programmes is often quite high as they are seeking to produce complex processes of behavioural, social, cultural and political change.

1.7 Duration: Even if the programme only has a relatively short duration, many of the outcomes only evolve over a relatively long period of time.

1.8 Does the programme challenge established systems enforcing unequal gender relations?: Probably most gender-focused programmes will challenge established systems to some extent, even if only in minor ways. But programmes with broader and more ambitious goals will usually rate high on this indicator.

Dimension 2: Institutions and stakeholders involved in the planning, financing, implementation and evaluation of the national programmes addressing the SDGs

In addition to the national governments that own and coordinate the implementation of the SDGs, most UN and official aid agencies are involved in most of the country programmes, as well as multiple international and national non-governmental organizations. There is a major challenge of coordination among all of these agencies. Past experience suggests that an additional challenge for evaluation is that much of the information on programmes is likely to not be well documented and monitored. In many cases, it may even be difficult to identify the programmes that are being implemented at the community and local level. Consequently, the overall complexity rating is again likely to be high:

2.1 Clarity of programme budgets: Some programmes are funded through general budget support, so it will often be difficult to track how funds from a particular donor were utilized or the total investment in each of the SDGs.

2.2 The number of funding and implementing agencies: This will vary significantly among countries, but in many cases, the number can be very large. Even in countries with fewer and small programmes, the number can be quite large.

2.3 The number of stakeholders: One of the goals of the SDGs is to increase the number of stakeholders by giving voice to vulnerable and previously excluded groups. The number is likely to be large and diverse.

2.4 Does the programme give voice to new stakeholders representing women and disadvantaged groups?: Gender-focused programmes frequently seek to give greater representation and voice to groups of women who are traditionally excluded from many consultations. As many of these groups have to be reached through non-traditional channels, their involvement often increases the complexity of the consultation process.

Dimension 3: The context with which the SDGs operate

National programmes addressing SDGs are designed and implemented in local, regional, national and international contexts in which they can be influenced by a multitude of economic, political, socio-cultural, environmental, historical and other factors. Each programme operates in a different context and is influenced by a different set of factors so the contextual analysis must be programme specific. An important element of a programme is its robustness and ability to operate as planned in the face of multiple and changing contextual factors. One of the goals of many gender programmes is to address many of the cultural, political and other barriers to gender equality and women's empowerment, and to promote changes in attitudes and behaviour so that what were once considered unacceptable or controversial programmes become mainstreamed and widely accepted.

3.1 Context dependence: The goal for many programmes is to reduce context dependency so that they are able to operate as planned without being forced to change in the face of the contextual factors. On the other hand, a goal of many programmes addressing SDGs is to increase their openness to the needs and suggestions of local communities, so in this case, the goal may be to increase responsiveness to certain kinds of social pressures. Consequently, judgment is required to assess the meaning of high and low ratings on context dependence.

3.2 Strength of contextual factors: As gender relations are affected by multiple mechanisms that seek to maintain the status quo, gender-focused programmes are likely to be affected by a wider range of contextual factors than many other non-gender programmes.

3.3 Is the programme likely to be affected by norms and power dynamics affecting the status of women?: Many SDGs will rate high on this indicator.

Dimension 4: Causality and change

As SDGs become larger and more complex, it becomes more difficult to assess what effects interventions have had on the intended programme outcomes. For example, programmes related to SDG-5 might include gender-awareness programmes for teenage girls through social media. One of the goals is to strengthen their empowerment and their confidence to combat negative gender images or behaviours of boys and other community members. However, there will often be a number of other programmes seeking to promote girls empowerment,

as well as other social, economic and political changes that affect the situation of young women. As programmes increase in scale and ambition, tracking causal changes becomes increasingly difficult:

4.1 Causal pathways: While causal pathways may be relatively simple and easy to track for small programmes with few components, tracking becomes increasingly difficult when there are multiple programmes with similar objectives. While approaches such as contribution analysis are a useful starting point, as causal pathways multiply in number and complexity (e.g. non-linearity and the same outcome being caused by different combinations of inputs and intervening variables), it often becomes necessary to draw on complexity science techniques such as systems mapping and social network analysis.[36]

In recent years, with the increasing availability of big data there has also been increasing interest in potential applications of smart data analytics to model causal pathways, particularly through the application of Baysian probability theory.[37]

4.2 Agreement on appropriate actions to address problems: In the health field, there will sometimes be a high degree of consensus on the best way to treat, for example, water-borne diseases, or in agriculture, there may be agreement on ways to increase crop-yield under certain conditions. However, in many areas there will be much less agreement.

4.3 Certainty on outcomes: For well-established programmes (for example, school meal programmes designed to increase school attendance), there may be a high degree of certainty on the likelihood that intended outcomes will be achieved. Whereas for new or complex programmes (for example, promoting women entrepreneurs in non-traditional sectors), the level of certainty may be much lower.

4.4 Complex process of gender-related behavioural change: Many of the processes of behavioural change affecting gender-relations and the status of women are subtle and difficult to capture. In addition, these processes operate at many different levels: interpersonal, within a multi-generational extended households, and within the

36 Bamberger, M., Vaessen, J., and E. Raimondo. 2016. *Dealing with Complexity in Development Evaluation: A Practical Approach*. Thousand Oaks, CA: Sage.

37 See Bamberger, M. 2016. "Integrating big data into the monitoring and evaluation of development programmes". *UN Global Pulse* (scheduled for publication in October 2016) for an introduction to the application of smart data analytics in the evaluation of development programmes.

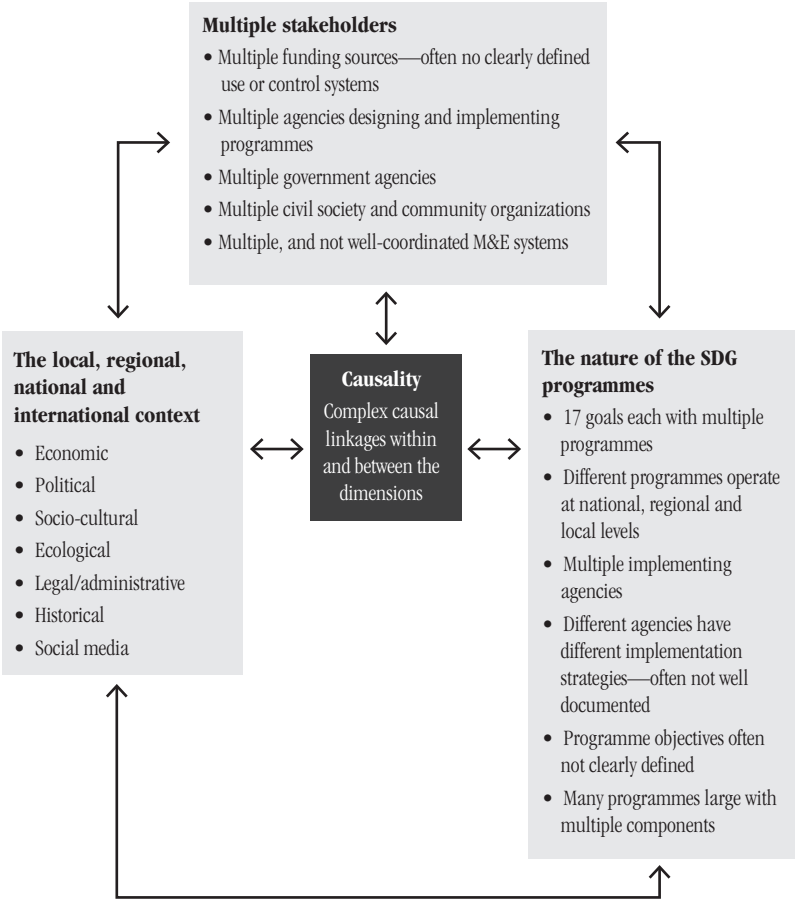
community, school, workplace, religious organization and through the mass media. Consequently, complexity is likely to be high on this indicator.

Applying complexity analysis to equity-focused and gender-responsive evaluation

All of the dimensions of complexity discussed in the previous section apply to equity-focused and gender-responsive evaluation. Due to the wide range of social mechanisms (social, economic, political, demographic, etc.) that affect and constrain the process of promoting gender and reducing inequalities, when resources permit, many programmes try to include a number of components to address several of these factors. Consequently, the nature of the intervention is likely to be complex. Similarly, there are likely to be a wide range of contextual factors that influence outcomes. The institutional dimension can vary in complexity depending on the number of funding and implementing agencies, but complexity will often arise from the need to involve a number of different agencies working on educational, economic, legal, infrastructure and community participation and empowerment. These are often agencies that do not normally work together leading to higher levels of complexity. Causality is also likely to be complex due to the subtle behavioural changes that are often difficult to define and capture. With respect to the nature of the evaluation, gender analysis often requires comprehensive statistical data on gender that are often not available and hinders efforts to advance gender equality.

Section 3.1 argued that inequality is caused by the interaction among multiple causes of unequal access to: resources and opportunities in the environment in which people grew up, and multiple social mechanisms in the context in which people live and work. The interactions among all of these factors determine whether or not observed differences in outcomes are judged to be fair and socially acceptable (resulting from free choice or unavoidable circumstances) or are unfair and socially unacceptable (resulting from preventable causes). The interactions among these multiple factors are likely to result in relatively high levels of complexity on all of the dimensions identified in Figure 1.

Figure 1. The SDG complexity framework



4. Proposed equity-focused and gender-responsive evaluation framework for SDGs

This chapter is based on the following three publications. For additional information and details on EFGR evaluation, we encourage you to consult them. They are all available on the Internet at http://www.evalpartners.org/library/selected-books and http://genderevaluation.unwomen.org/en/evaluation-handbook free of charge.

- **Bamberger and Segone. 2011.** "How to design and manage equity focused evaluations". New York, NY: UNICEF.
- **UN Evaluation Group. 2013.** "Integrating human rights and gender equality in evaluation: Guidance document".
- **UN Women. 2015.** "How to manage gender-responsive evaluation: Evaluation handbook". Independent Evaluation Office. New York, NY: UN Women.

In addition, we invite you to take the related free e-learning programme available at: http://elearning.evalpartners.org/.

The guidance presented in this report provides an initial framework for country led evaluations that are equity focused and gender responsive, and that will evolve in parallel with refinements to the overall SDG follow-up and review mechanism. They will also be adapted to the priorities and capacities of individual countries. The approach builds on the central SDG goal of "No-one left behind".

Other key features of the proposed approach are that it is country-led and based on a participatory, consultative process involving a broad range of stakeholders at the national and local levels. The approach combines the collection of a basic set of indicators that are broadly comparable across countries and regions, with maximum flexibility to reflect country differences in the approaches to equality and gender dynamics and to respond to national priorities and research capacities. The proposed approach also recognizes that there is no single, "best" evaluation methodology and that different countries and organizations can adopt different approaches. In particular, the need is recognized to combine quantitative indicators that permit the measurement of broad trends and cross-country comparisons (reflecting the fact that the SDGs must adopt a global, holistic focus that recognizes interactions among countries and sectors), with more in-depth qualitative measures that reflect the complex, multi-dimensional and context specific nature of gender and reducing inequality. A challenge for the evaluation of the SDGs is to integrate the quantitative and qualitative

approaches and to recognize the different purposes of each and how they complement each other. Neither approach is better or more "rigorous", and each has its own specific standards of rigor.

4.1 The proposed country-level, equity-focused and gender-responsive evaluation framework

All countries have committed to conduct national SDG reviews, although the focus and depth will vary according to country priorities, data collection and analysis capacity. The EFGR evaluations should form an integral part of the national SDG review mechanisms. Both concepts of gender and reducing inequalities are interpreted differently in different regions and countries, so the focus and scope of the EFGR evaluation may vary depending on how these concepts are interpreted in each country.

In the UN Women Evaluation Handbook on "How to manage gender-responsive evaluation",[38] gender-responsive evaluation is defined as having two essential elements: what the evaluation examines and how it is undertaken. Gender-responsive evaluation assesses the degree to which gender and power relationships—including structural and other causes that give rise to inequalities, discrimination and unfair power relations, change as a result of an intervention using a process that is inclusive, participatory and respectful of all stakeholders (right holders and duty bearers). Gender-responsive evaluation promotes accountability to gender equality, human rights and women's empowerment commitments by providing information on the way in which development programmes are affecting women and men differently and contributing towards the achievement of these commitments. It is applicable to all types of development programming, not just gender-specific work.[39]

UNICEF defines equity-focused evaluation as a judgment of the relevance, effectiveness, efficiency, impact and sustainability of policies, programmes and projects concerned with achieving equitable development results[40](see Box 6). In humanitarian evaluations, coverage, connectedness and coherence are also addressed. The approach involves rigorous, systematic and objective processes in the design, analysis and interpretation of information in order to answer specific

38 UN Women. 2015. "How to manage gender-responsive evaluation: Evaluation handbook". Independent Evaluation Office. New York, NY: UN Women.

39 UN Women undated.

40 Bamberger and Segone, 2011.

BOX 6. Examples of definitions of equity-focused and gender-responsive evaluations

Equity-focused evaluation
UNICEF defines equity-focused evaluations as "an assessment made of the relevance, effectiveness, efficiency, impact and sustainability of interventions on equitable development results." Equity-focused evaluations look at structural bottlenecks and power relationships, and use an empowerment process.

Gender-responsive evaluation
UN Women subscribes to the UNEG definition of evaluation but directly incorporates principles of gender equality, women's rights and the empowerment of women. UN Women defines evaluation as "a systematic and impartial assessment that provides credible and reliable evidence-based information about the extent to which an intervention has resulted in progress (or the lack thereof) towards intended and/or unintended results regarding gender equality and the empowerment of women. As a process itself, evaluation is also a means to enhance gender equality and the empowerment of women through the incorporation of gender and women's rights dimensions into evaluation approaches, methods, process and use. Accordingly, not only does evaluation act as an important driver of positive change towards gender equality and the empowerment of women, but the way in which the evaluation process itself is undertaken empowers the stakeholders involved."

Sources: Bamberger, M. and M. Segone. 2011. "How to design and manage equity-focused evaluations". UNICEF. UN Women Evaluation Policy, 2013.

questions, including those of concern to worst-off groups. It provides assessment of what works and what does not work to reduce inequality, 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.[41]

The UNEG Guidance Document "Integrating human rights and gender equality in evaluations" provides a valuable resource for all stages of the formulation, design, implementation, dissemination and use of EFGR-focused evaluations.[42]

EFGR evaluations are conducted in a participatory, consultative way with the active involvement of government agencies, civil society, and other national stakeholders as well as the country offices of UN and donor agencies. While there will be considerable variation among countries, the evaluations are likely to follow, with some adjustment to each country context, the stages described below.

41 Bamberger and Segone, 2011, p. 9.

42 UN Evaluation Group. 2013. "Integrating human rights and gender equality in evaluations: Guidance document."

Stage 1: Broad-based consultations to select and develop the equity-focused and gender-responsive evaluation design

The following are some of the questions to be defined during the consultations:

a) Defining the key evaluation questions

Before considering whether or not additional indicators are required or what the appropriate evaluation designs are, it is essential to agree on what the key questions are that the evaluations must address. While this may seem obvious, many evaluators are concerned to propose their preferred methodologies and often jump straight into methodological debates before having clarified the questions that must be addressed. It is important that evaluation questions be defined in consultation with stakeholders and not by the evaluation specialists.

Chapter 1 identified three levels at which evaluations can be conducted (policy, programme and project), and four types of evaluation (policy, formative, developmental and summative). Each type of evaluation is designed to address different kinds of questions (see Table 1 in Chapter 1) which are of interest to different stakeholders. It is essential to identify the kinds of information that each stakeholder group requires before selecting the appropriate kinds of evaluation design. It is very likely that the evaluation of the EFGR dimensions of the country SDG programmes will require conducting evaluations at different levels and using several different evaluation types at different points in the evaluation.

For EFGR evaluations, it is important to ensure that evaluation questions for most if not all of the SDGs specifically identify and address relevant gender and equality issues. In order to achieve the objective "No one left behind", issues of gender equality and reducing inequalities must be addressed throughout the SDG evaluation, as most interventions have differential impacts on women and men, and there is always the challenge of identifying groups who may be left out and who do not benefit from the different programmes.

Table 3 illustrates some of the evaluation designs that can be used to address the different summative evaluation questions. This table is limited to summative evaluation both for reasons of space and because summative evaluations can use all of the different evaluation methods that are used for the other three kinds of evaluation.

Table 3. Impact (summative) evaluation questions and possible evaluation designs

Key evaluation question	*Possible evaluation designs**
1. To what extent can a specific outcome (or impact) on gender or on reducing inequalities be attributed to the intervention?	• Experimental and quasi-experimental designs • Statistical designs • Hybrid case studies and participatory designs
2. Has the intervention made a difference to intended or unintended outcomes on gender or outcomes related to reducing inequalities?	• Experimental and quasi-experimental designs • Contribution analysis • Qualitative comparative analysis (QCA) case study methods
3. How has the intervention made a difference to outcomes on gender or to outcomes related to reducing inequalities?	• Theory-based methods • Realist evaluation • Participatory approaches • Theory of change • Process tracing
4. Can this be expected to work elsewhere with respect to outcomes on gender and outcomes on reducing inequalities?	• Participatory approaches • Natural experiments • Review and synthesis studies • QCA case studies

* The designs are described in Section 4.2.

Source: Adapted from Stern et al. 2012, and Bamberger et al. 2016.

b) Selecting the best evaluation design (or combination of designs) to address each evaluation question

Table 3 lists some of the evaluation designs that can be used to address each of the four evaluation questions. These designs are discussed below in Section 4.2. All of the designs should incorporate the EFGR approach and all designs should include questions relating to gender and reducing inequalities.

c) What are the levels at which the EFGR evaluations will be conducted?

Evaluations can be conducted at the national, sector (programme) or local levels, and for some large countries, at the state, provincial or regional levels. The choice will depend on country priorities and resource availability.

d) Is there an integrated country evaluation strategy?

Ideally, all evaluations should be selected as part of an integrated SDG evaluation strategy (and responding to the key evaluation questions), but often individual evaluations will be selected in an ad hoc

manner as resources become available or depending on the interest of different agencies.

Stage 2: Selecting the best combination of evaluation designs

There are six main evaluation designs that can be used in the EFGR evaluations. No single design can address all of the evaluation questions, so normally a combination of several designs may be used. In cases where the programme is considered complex, there are a number of additional complexity-responsive evaluation designs that can be used. The designs are described in Section 4.4. It is important to recognize that there is no single "best" evaluation design and that the choice of design will be determined by the questions being asked, the purposes of the evaluation and the nature of the programme ("evaluand") being studied.

Stage 3: Complexity-responsive approaches

When programmes are defined as complex (see Chapter 3 and Section 4.4), conventional evaluation designs will usually not be able to assess the effectiveness of a particular project in contributing to desired changes in each of a number of outcomes that are produced through complex causal chains. In these circumstances, it will often be necessary to consider the use of complexity-responsive evaluation designs. These are discussed in Section 4.4.

Stage 4: Evaluating resilience and sustainability

A central goal of the SDGs is to promote resilient programmes and societies and to ensure that the benefits of the programmes are sustainable over long periods of time. As most conventional evaluations do not assess the long-term sustainability and resilience of programmes, these questions require the use of special evaluation tools and techniques that are described in Section 4.5. Special issues and challenges affecting the sustainability and resilience of programmes for gender and for reducing inequalities and goals must also be assessed.

Stage 5: Conducting an evaluability assessment

Once the methodologies have been developed, it is important to conduct an *evaluability assessment* to ensure that the proposed evaluation designs can adequately address the key EFGR questions and that this can be done within the time, budget, human resources and institutional constraints of the agencies conducting and managing the evaluations.

Collection and analysis of country-level indicators in the basic SDG indicator set will be a key element of the evaluability assessment. Each country will collect and analyse as many as possible of the recommended SDG basic indicator set. For the purposes of the EFGR evaluation, particular attention will be given to the indicators on gender and on reducing inequalities.

The data for the indicators to measure the different SDG goals and targets will be collected through a wide range of agencies.

- Exploratory data analysis and data classification. This will examine the coverage, implementation, completeness and quality of the SDG basic indicators.
- Analysis of the indicators for gender and for reducing inequalities. The analysis will cover both the SDG-5 and SDG-10 sub-indicators as stand-alone goals as well as the gender dimensions of other SDGs. UN Women will issue a monitoring report in November 2017 that will provide a comprehensive and authoritative assessment of progress, gaps and key challenges in the implementation of the SDGs from a gender perspective. The report will be based on the latest quantitative analysis of global and regional trends in achieving the SDGs for women and girls as well as in depth-qualitative country studies that examine how the SDGs are being implemented at the national level.
- Addressing "No-one left behind". Available methodologies will be used[43] and, where necessary, new ones will be developed to identify groups that are excluded from benefits based on gender, income, age, ethnicity, physical or mental disadvantage, geographical location or other factors.

At this point, there may be a discussion on whether or not additional indicators and information should be included. The basic SDG indicators are not intended to be comprehensive due to the need to keep the list relatively short to make it manageable. Consequently, there are many areas in which additional indicators for gender and for reducing inequalities could be identified.

Once the evaluation is launched, there will be periodic meetings to review progress and to make any required adjustments to the design. A quality control system should be put in place to ensure the quality and completeness of the data.

43 Bamberger and Segone 2011.

Stage 6: Special data collection and analysis tools for equity-focused and gender-responsive evaluations

EFGR evaluations use data collection and analysis tools described in this chapter. However, these must often be complemented by special data collection techniques to address issues relating to:

- The disaggregation of household-level data to examine how resources are distributed among different household members. An example of a widely used tool is the Harvard Gender Analysis Framework, which provides a checklist for examining time-use and access to and control of productive resources. Nutrition studies have also developed very detailed tools to monitor in detail questions such as food consumption by each household member.
- Capturing information on sensitive topics such as household decision-making, domestic violence, family planning and sexual behaviour that women and other household members are reluctant to discuss with interviewers. A range of participant observation, key informant, focus group and qualitative techniques are available for this purpose.
- Monitoring behavioural change. Intended outcomes of gender interventions often include behavioural changes that are difficult to observe. Respondents are often not aware of these subtle changes (for example, in decision-making, resource control or women's feeling of self-confidence). Participant observation, where people are observed over time, can often be used for this purpose.
- Analysis of the multiple social control mechanisms that constrain women's personal freedom, access to resources and participation in decision-making: The complexity-responsive tools and approaches for this are discussed later in this chapter.
- Case studies provide useful tools to complement all of the above as they can document changes over time, the complex processes affecting women's access to services, and their ability to participate in social organizations and decision-making.

Stage 7: Dissemination of equity-focused and gender-responsive evaluations

There should be a programme of dissemination that combines conventional written reports with more creative methods in coordination with mass media and social media (including the use of new information technology such as mobile phones, tablets and social media where appropriate) and through workshops and other kinds of briefings. It is

also important to ensure that findings are accessible to local communities and, where appropriate, theatre, dance and other culturally appropriate communication methods may be used. The findings must also be available in local languages.

4.2 Main types of evaluation designs that can be considered for equity-focused and gender-responsive evaluations

There are six widely used evaluation designs, all of which could be applied in the EFGR evaluation. The designs (summarized in Table 4) are the following:

Experimental and quasi-experimental designs

These designs involve the comparison of a treatment (project) and a matched control group that does not have access to the project (and which helps define the counterfactual). This design is most widely used for project evaluation as there is usually a simple linear causal relation that can be easily measured, but similar designs can often be used at the programme level. However, the greater difficulty of defining a comparison group means that quasi-experimental designs are more commonly used at the programme level. Project impact is estimated by comparing the change in the outcome variable for the two groups between the start of the project (baseline) and the project completion. There is an important distinction between experimental designs. The most common are: RCTs, where subjects can be randomly allocated to the project and control groups, thus eliminating differences between the two groups that might explain the outcomes; and quasi-experimental designs, where subjects cannot be randomly assigned. In this latter case, a matching technique must be used to select a comparison group that is as similar as possible to the project group. Quasi-experimental designs are subject to the problems of *selection bias,* as differences in outcomes, which are interpreted as showing project effects, may in fact be due to differences between the two groups, which the matching procedure was not able to control. Quasi-experimental designs are very widely used, and there are many different variations that vary in terms of their methodological rigor.

Over the past few years, there has been a significant increase in the application of RCTs, including in the field of evaluation of gender programmes. A practical limitation of these designs is that they tend to be quite expensive and require a high level of technical expertise.

Recently, there are some promising developments for RCT and quasi-experimental design evaluations conducted on mobile phones and related information and communications technologies, where costs can be significantly reduced. These have not been used widely for EFGR evaluation, partly because of the difficulty of a precise identification of the vulnerable groups.

Similarly experimental designs have not been used extensively for EFGR evaluation, partly because of the difficulty of a precise identification of the vulnerable groups and of the process of change. Outcomes on gender and reducing inequalities usually involve multiple processes of behavioural change, which are difficult to fit into the precisely defined linear process of RCTs. However, over the past few years, there has been a steady increase in the application of experimental designs to assess gender interventions that have a single or only a few inputs and intended outcomes.[44]

Table 4. The main evaluation approaches that can be used in the EFGR evaluations

Design	*Variations*	*Examples and references*
1. Experimental and quasi-experimental	RCTs, quasi-experimental designs, natural experiments	• Using RCT to evaluate the impacts of training of cross-border guards in Rwanda to reduce violence against women and improve socio-economic outcomes for women (Source: World Bank Gender Innovation Lab) • Many of the RCTs conducted by the Poverty Action Lab assess the impact of development interventions on women (www.povertyactionlab.org)
2. Statistical	Statistical modelling, econometrics, public expenditure incidence analysis, public expenditure tracking	• Public expenditure incidence analysis used to assess what proportion of public expenditures in sectors such as health and education go to low income families (Source: IMF)
3. Theory-based	Theory of change, process tracing, contribution analysis, realist evaluation	• Using theory of change and contribution analysis to assess the effectiveness of a 10-year programme to reduce violence against women in El Salvador (Source: OXFAM USA)

44 The World Bank Africa Gender Innovation Lab has conducted experimental and quasi-experimental designs to evaluate gender interventions in the areas of land titling, agriculture, private sector development and youth employment. http://www.worldbank.org/en/programs/africa-gender-innovation-lab

4. Case-based	Naturalistic, grounded theory, ethnography, process tracing, QCA, within-case analysis, simulations, network analysis	• QCA used to assess the effectiveness of UN Women interventions at the national level on women's economic empowerment; the country was used as the unit of analysis (Source: UN Women).
5. Participatory and qualitative	Participatory and empowerment evaluation, feminist evaluation, participatory appraisal methods, most significant change, outcome harvesting, outcome mapping	• Village women design a survey instrument to identify family needs in poor communities in India and then interpret and disseminate the findings (Source: World Bank Social Observatory, India) • Evaluating gender structural change: the experience of the GENOVATE project's evaluation (Source: María Bustelo, Julia Espinosa and María Velasco, Complutense University of Madrid)
6. Review and synthesis	Meta-analysis, narrative synthesis, realist synthesis	• Using a systematic review, covering all of the published literature, to assess the impact of micro-credit on women's economic empowerment (Source: Vaessen, Rivas and Leeuw, 2016)

Sources: Adapted from Stern et al. 2012, and Bamberger et al. 2016.

Statistical designs

These use statistical modeling and econometric analysis and are mainly used at the national level or for cross-country comparisons. A typical example would be to use a cross-country comparison to assess the effects of a particular policy intervention on, for example, the supply of low-income housing while controlling for national level indicators such as GDP, average educational level, urban growth rates and unemployment. To date, statistical designs have not been widely used in the evaluation of gender policies and interventions.

There is extensive economic research literature on who is reached by different kinds of public expenditure, particularly for health, education and conditional cash transfer payments. These studies use national income and expenditure data, so the analysis is limited by the fact that available data frequently only permits disaggregation by income, often not even by sex. However, they do provide a useful first approximation on which income groups do and do not have access to different services.

A potentially powerful statistical tool for EFGR evaluation is the social exclusion approach. This examines how exclusion results from the interaction among different dimensions such as age, sex, income,

ethnicity and geographical location.[45] While only feasible in countries with good national data sets, it is an attractive approach as it avoids the sector-by-sector analysis of equality that is often done.

Theory-based evaluations

These include widely used designs such as theory of change, contribution analysis and process tracing.[46] With all of these designs, an *intervention theory* identifies a set of causal assumptions explaining how an intervention is intended to work, the causal linkages between different levels and components of the programme, and the critical assumptions that must be tested. The theory of change helps guide the evaluation design the evaluation questions, assumptions and the definition of indicators and provides a framework for the interpretation of the results. Theory-based approaches have been widely applied in gender-responsive evaluations.[47]

A widely used variant is *contribution analysis,* which is used to assess the contribution of a particular programme in complex situations where it is not possible to use experimental designs and quasi-experimental designs.

Theories of change are used in almost any evaluation context and can vary from very simple one-page models to complex, software-driven models explaining the interactions among multiple components and on multiple levels. Theories of change have a major advantage in that they are used in a participatory way. User friendly online software also makes it possible to involve stakeholders who are unable to physically meet. It is particularly important when developing a gender-responsive theory of change to ensure that this is developed in a bottom-up participatory way with all women's groups being actively involved in the process.

---

45 The social exclusion approach originated in Europe to assess factors affecting the access of migrant populations to public social services. For an example of these approaches, see the 2015 Annual Report of the Centre for the Analysis of Social Exclusion at the London School of Economics. http://sticerd.lse.ac.uk/dps/case/cr/casereport98.pdf. For an example of how social exclusion is now being applied in developing countries, see the work of World Health Organization. http://www.who.int/social_determinants/themes/socialexclusion/en/

46 Leeuw, F.L. 2012. "Linking theory-based evaluation and contribution analysis: Three problems and a few solutions". *Evaluation*, July 2012: 18 (3): 348-363.

47 See for example, UN Women. 2014. "An empowered future: Corporate evaluation of UN Women's contribution to women's economic empowerment". Independent Evaluation Office. New York, NY: UN Women, and the ActionAid theory of change for tackling violence against women: http://www.actionaid.org.uk/sites/default/files/doc_lib/theory_of_change_on_vawg.pdf Also UNFPA. 2016: Evaluation of UNFPA support to population and housing census data to inform decision-making and policy formulation (2005-2014). Evaluation Office.

A potentially powerful theory-based approach is the bottleneck analysis developed by UNICEF.[48] This examines how access to, or exclusion from, services such as health and education is determined by the interactions among supply-side, demand-side and contextual factors and how services are actually used by worse-off groups. It is used to assess access of all vulnerable groups or it can focus specifically on factors affecting the differential access of women and men to different services.

Case-based approaches

There are a wide variety of case-based evaluation methods, some of which are largely descriptive while others permit quantitative analysis. The common elements are that they all take the case (the individual, household, group, community, organization, etc.), rather than a single variable, as the unit of analysis. Each case is considered unique, with different combinations of factors so that the analysis looks for diversity rather than homogeneity. The two most common case methods are descriptive case methods[49] that are familiar to all evaluators and qualitative comparative analysis (QCA).[50] QCA refers to a family of methods that focus on a limited number of empirical cases, for which configuration of effects (outcomes, impacts) and conditions for effects to occur are explored. The analysis identifies the set of attributes (characteristics) that must be present for the outcome to occur (the necessary conditions) and the set of attributes that are always present when the outcome does not occur. The evaluation of UN Women's contribution to women's economic empowerment is an example of a QCA analysis where the cases represent countries where UN Women programmes are operating.[51]

Descriptive cases are used for in-depth understanding of how different groups actually experience social exclusion, and QCA can be used to understand the configuration of factors that promote or overcome exclusion.

Descriptive case study designs are used very widely in gender-responsive evaluations.

48 Bamberger and Segone 2011, pp. 45-50.

49 Yin, R. 2012. *Applications of Case Study Research.* Thousand Oaks, CA: Sage.

50 Byrne, D. and C. Ragin. 2009. *The Sage Handbook of Case-Based Methods.* Thousand Oaks, CA: Sage.

51 UN Women. 2014. "An empowered future: Corporate evaluation of UN Women's contribution to women's economic empowerment". Independent Evaluation Office. New York, NY: UN Women.

Case studies have the flexibility to describe in-depth the lived experience of women and how they interact with family, community members and outside groups. It is also possible to capture processes and social pressures that are more difficult to capture in surveys. International agencies often select country case studies when they are assessing the effectiveness of their global or regional programmes. The challenge with case studies is to ensure that they are selected in a way that ensures the sample is broadly representative of the groups being studied so that it is possible to generalize from the findings.

Over the past few years, there has been growing interest in QCA methods. These methods have the advantage that they can focus on both the unique features of each case (individual, household, organizations) as well as the combination of factors that determine the presence of absence of desired outcomes. The approach is helpful for understanding interactions among the multiple factors that affect outcomes.

Participatory and qualitative designs

"Participatory evaluation designs involve a wide range of stakeholders in the design, implementation, and interpretation of the evaluation. Participatory methods may be used for methodological reasons, to strengthen data quality and validity or for ideological reasons."[52]

Participatory methods often use a mixed methods approach where the findings from different data collection and analysis methods are triangulated to increase validity. Many participatory methods are also used in empowerment, gender-responsive or equity focused evaluations as part of a process of political or social empowerment.[53] Participatory approaches are in-line with the principles of human rights and gender-responsive evaluations (see, for example, the UNEG Guidance).

Some of the widely used participatory and qualitative evaluation methods include:

- Outcome mapping[54]

52 Cousins, J. and E. Whitmore. 1998. *Framing Participatory Evaluation: New Directions for Evaluation.* Wiley Periodicals.; Bamberger, M., Vaessen, J., and E. Raimondo. 2016. *Dealing with Complexity in Development Evaluation: A Practical Approach.* Thousand Oaks, CA: Sage.

53 Mertens, D. M. and S. Hesse-Biber. 2013. "Mixed methods and credibility of evidence in evaluation". *New Directions for Evaluation,* 2013: 5–13. doi: 10.1002/ev.20053.

54 Earle, S., Carden, F. and T. Smytlo. 2001. "Outcome mapping: Building learning and reflection into development programmes". Ottawa: International Development Research Center.

- Outcome harvesting and other story-telling methods[55]
- Most significant change[56]
- Participatory appraisal methods[57]
- Key informant interviews
- Some applications of focus groups
- Participant observation and longitudinal case studies[58]

Participatory and qualitative methods are used widely in EFGR evaluation and are in line with human rights and gender-responsive approaches. Participatory appraisal techniques have been used whereby communities discuss and identify mechanisms of exclusion. Sometimes this is done through social maps constructed by the community indicating who does and does not have access to different resources.

Participatory approaches are also very useful for conducting rapid exploratory and diagnostic studies in order to fully understand the context in which the programme operates before deciding on the evaluation methodology.

Participatory methods involve target populations and vulnerable groups in the definition of the purpose and methodology of the evaluation, the collection of data and the interpretation of findings. These methods are well suited to evaluations designed to promote empowerment and social justice and, in particular, to ensure that all women are involved in the process.

Focus groups are one of the widely used participatory methods, and when they are well facilitated, they can ensure that all members of the group have the opportunity to speak. Many gender researchers believe that focus groups provide a supportive environment for less vocal women to express their views as they are not dominated by the interviewer, which can happen when one-on-one interviews are conducted.

Participatory methods are often combined with theory of change techniques, such as most significant change, outcome mapping and

55 Wilson-Grau, R., and H. Britt. 2012. "Outcome harvesting". Cairo, Egypt: The Ford Foundation's Middle East and North Africa Office. Available at: http://www.outcomemapping.ca/resource/resource.php?id=374.

56 Davies, R. and J. Dart. 2005. "The 'most significant change' (MSC) technique: A guide to its use". Melbourne. Available at http://www.mande.co.uk/docs/MSCGuide.pdf.

57 Kumar, S. 2013. "Understanding equity and growth from a development perspective: Background paper". New York, NY: Rockefeller Foundation.

58 Salmen, L. 1987. *Listen to the People: Participant Observer Evaluation of Development Programmes*. New York, NY: Oxford University Press.; Hashemi, S., Schuler, S., and A. Riley. 1996. "Rural credit programmes and women's empowerment in Bangladesh." *World Development* 24(4) 635-53.

outcome harvesting, all of which seek to enable local communities to express their opinions on what should be, or what were the, effects of the interventions.

Review and synthesis

These approaches involve the identification of all evaluations that have been published on a particular topic and that satisfy certain methodology quality standards. The findings are synthesized to highlight the main findings and lessons.[59] These reviews provide a wider perspective and help users understand how implementation and outcomes are affected by different local contexts. There are a number of agencies that now regularly publish these reviews.[60]

4.3 Some general strategies for integrating equity-focused and gender-responsive principles into SDG evaluations

With time, all SDG evaluations should be EFGR. However, in recognition that in some countries this will be an incremental process, there are a number of general approaches that can be helpful for integrating considerations regarding gender and reducing inequalities into the SDG evaluations. In addition to the general evaluation methodologies discussed above, this section presents a number of approaches that can be helpful for integrating gender and equity considerations into the SDG evaluations (Box 7).

Review of lessons learned from past approaches

Many agencies have found it useful to conduct a review or evaluation of past gender mainstreaming policies, strategies and approaches and the lessons learned from these. This is also a way for making the case for greater institutional change beyond the evaluation function.

59 Vaessen, R. and M. Bamberger. 2016. "Impact evaluation approaches and complexity" in Bamberger, M., Vaessen, J., and E. Raimondo. 2016. *Dealing with Complexity in Development Evaluation: A Practical Approach*. Thousand Oaks, CA: Sage.

60 The International Initiative for Impact Evaluation is one of the best known in the development field (3ieimpact.org), which now has information on more than 2,500 impact evaluations. Many UN agencies and bilaterals also commission systematic reviews from time to time.

Building gender equality into the theory of change and results framework of programmes

This should be done as both a stand-alone outcome and integrated/ mainstreamed throughout.

BOX 7. Lessons from evaluation practitioners on promising tools for equity focused and gender responsive evaluations

During January and February 2016, online consultations were conducted with evaluation practitioners from around the world to solicit their experience and recommendations on the evaluation of the SDGs with an EFGR lens. Similar consultations were conducted on evaluation of complexity, national evacuation systems, and factors affecting the utilization of evaluations (see Annex 3). All the discussions are available online at: http://gendereval.ning.com/forum/topics/eval-sdgs.

This box includes a summary of feedback on useful evaluation tools.

With respect to promising evaluation tools and approaches, the focus was on participatory methods stressing the importance of broad-based consultations with local communities at all stages of the evaluation process. The following are some of the suggestions and tools that were found useful:

- It is important to institutionalize and systematize data collection
- Oral histories and ethnographic tools
- Evaluations must be owned by the community
- Mixed methods are useful but often they cannot dig deep enough into "sticky" gender issues where women may be reluctant to respond
- Participatory tools
- Spatial mapping
- Gender profiling of public expenditures and gender-based auditing
- Institutional gender analysis
- Seasonal mobility mapping
- Strength-based community life approach
- Life course approach
- Community mobilizing to give voice to vulnerable groups
- Outcome mapping
- Participatory statistics
- Post-project sustainability analysis
- Power analysis among community groups and between the community and external groups

Developing a checklist of key areas where an evaluation can integrate gender perspective

Checklists of important issues to be analysed from a gender perspective have proved a useful reference tool in the design of both gender-responsive programmes and evaluations.[61] Checklists can be both general (covering all issues to be addressed in gender-responsive evaluations) or sector specific. However, they should avoid being prescriptive or becoming a box-ticking activity and rather be used as a means for incentivizing thinking on important aspects for consideration.

The Harvard Gender Analysis Framework includes two widely used checklists for the analysis of access to and control of productive resources, disaggregated by sex, and where required by age, civil status or otherwise relevant variables.[62]

Within the context of the UN System-wide Action Plan for gender equality and women's empowerment[63], which constitutes the first accountability framework for gender mainstreaming in the UN system, UNEG has developed a Scorecard[64] to support reporting on the Evaluation Performance Indicator. The Scorecard is meant to be used along with the UNEG Guidance on Integrating Human Rights and Gender Equality in evaluation, and provides a set of four criteria upon which UN entities assess the integration of gender equality in evaluation reports, which results in an overall rating for the evaluation report. The criteria of the scorecard focus on: a) integration of gender equality in evaluation scope, evaluation indicators, evaluation criteria and evaluation questions; b) gender-responsive methodology, methods, tools and data analysis techniques; and c) the evaluation findings, conclusions and recommendations reflect a gender analysis.

---

61 A number of organizations such as UN Women, USAID, DIFD and CARE international have developed checklists for staff to assess how well gender issues have been addressed in the design and implementation of their programmes. For example, DFID staff are asked questions such as: "Have we counted all women and girls?", "Have both women and men been consulted?", "Have we invested equally in women and men?", "Do women and girls receive a fair share (of programme resources?)"

62 Overholt, C., Anderson, M., Cloud, K., and J. Austin. 1985. *Gender Roles in Development Projects: A Case Book*. Connecticut: Kumarian.

63 To learn more about the UN System-wide Action Plan for Gender Equality and Women's Empowerment (UN-SWAP), visit the website: http://www.unwomen.org/en/how-we-work/un-system-coordination/promoting-un-accountability.

64 UNEG, UN-SWAP. "Evaluation Performance Indicator Technical Note and Scorecard". http://uneval.org/document/download/2148.

Rapid diagnostic studies

In cases where it will not be possible to conduct a full gender-responsive evaluation, a rapid diagnostic study can help identify issues that should be addressed in the evaluation. White recommends the use of "economic ethnography" whereby project managers review the existing ethnographic literature before designing a programme or evaluation. Gender analysis in programme design and implementation is also recommended as a way to strengthen the EFGR lens of the programme.[65]

Integrating gender into ongoing or planned evaluations

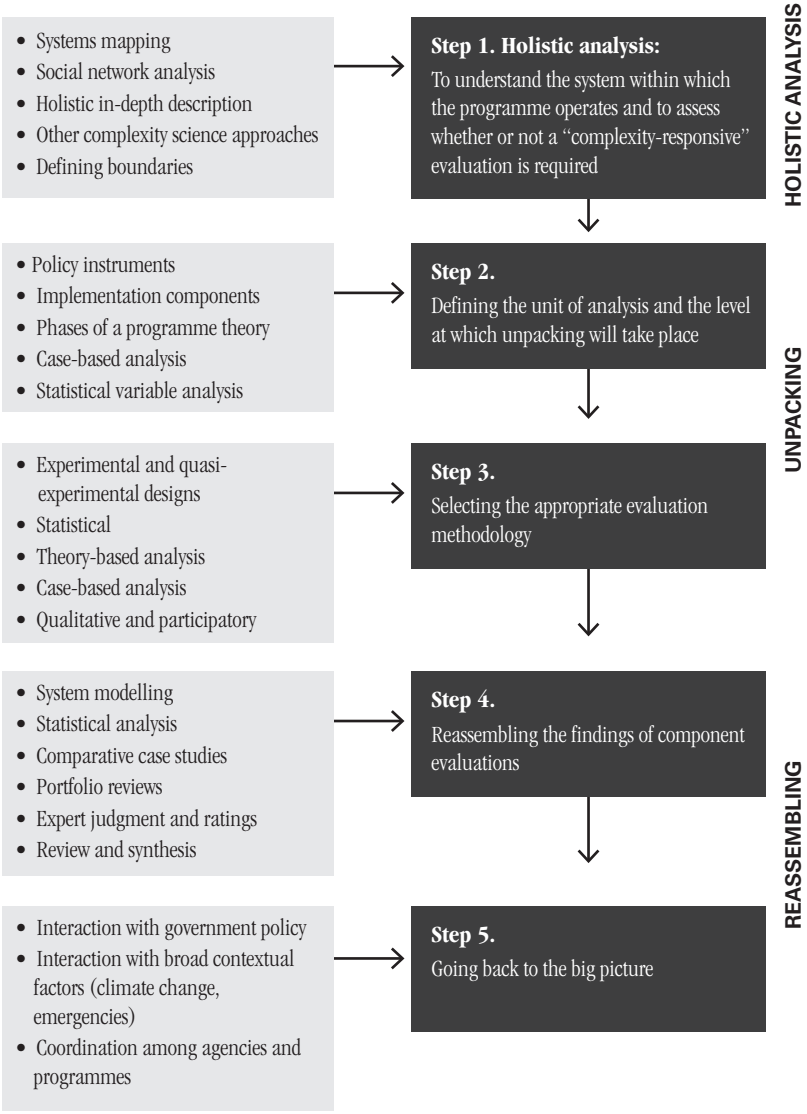
An economical way to conduct a gender-responsive evaluation is to take advantage of a study that is being planned by another agency. This "piggy-back" approach can range from requesting that key data is sex-disaggregated during the data collection and analysis process, through including additional questions in the survey, to developing a special module to be administered to a sub-sample of wives, female students or older women.

4.4 Approaches for the evaluation of complex development programmes

As development initiatives become more complex, conventional evaluation approaches are no longer able to fully evaluate how multiple interventions funded, designed and implemented by multiple stakeholders, and operating in complex environments, contribute to observed changes in multiple (intended and unintended) outcomes. Under these increasingly common scenarios, it becomes necessary to find new evaluation approaches that are "complexity-responsive" and equity-focused and gender responsive. The following five-step approach (see Figure 2) provides an initial framework for the evaluation of the many kinds of complexity programmes that are planned under the SDG initiative. This framework also applies to the evaluation of complex EFGR initiatives.

65 White, H. 2009. "Theory-based impact evaluation: Principles and practice". *Journal of Development Effectiveness,* 1 (3), 271-284.

Figure 2. A five-step approach for the evaluation of complex programmes



Step 1: Holistic analysis

The evaluation begins with a holistic analysis to understand multiple systems within which the programme operates. This builds on the four dimensions of complexity discussed in Chapter 3:

- The nature of the intervention
- The organizational framework and the interactions among stakeholders
- The contextual factors affecting the how the programme is designed, implemented and evaluated
- The complex causal pathways between inputs and outcomes

The concept of boundaries is important. Boundaries define how broad programme effects are intended to be and how widely effects will be assessed (these are two separate but related issues). For example, is a programme (such as a girl's scholarship programme) only designed to benefit girls and the target villages or districts, or it is designed to have spillover effects in surrounding areas? There are similar decisions to be made with respect to the evaluation. Is the evaluation only designed to assess direct effects or also spillover effects? Boundaries must also be assessed with respect to time horizons. Will effects only be assessed over one year or over longer periods of time?

The narrower the boundaries, the more economical and precise the evaluation. However, there is a trade-off, as potentially important secondary effects (both positive and negative) will not be captured.

Figure 2 identifies some of the complexity science tools that can be used in this holistic analysis. This analysis determines whether or not the programme can be considered sufficiently complex to justify a complexity-responsive evaluation, and what are the main elements of complexity that must be addressed in the evaluation. The checklist given in Chapter 2 (Table 2) illustrates a useful approach for rating the level of complexity of the four dimensions.

Step 2: Unpacking the programme into its main components

This is "unpacked" into its main components or elements, each of which can be evaluated separately. A big advantage of this approach is that it is possible to use conventional evaluation designs to evaluate the individual components (whereas these designs do not generally work to evaluate the whole complex programme).

Step 3: Selecting the appropriate evaluation methodology

All of the six conventional evaluation designs (discussed earlier in Chapter 4) can be used, where appropriate, for these evaluations.

Step 4: Reassembling the findings of the individual component evaluations

This is a very important phase, as there are many situations in which each individual component is evaluated positively, but the whole programme makes little contribution to its overall objectives. There are at least three reasons for this:

- The programme goal is over-ambitious: for example, a training and awareness-raising programme for women entrepreneurs may to be too small and limited to address the multiple social, economic, political, legal and cultural barriers to women's empowerment.
- There may be problems of coordination among the different agencies and programme components.
- Unforeseen events, such as a drought, civil war or change in international markets may seriously limit the programme's effects.

Figure 2 lists some of the methodologies that can be used for this reassembling analysis.

Step 5: Going back to the "big picture"

Finally, the programme and its effects must be assessed with the context of government policies and other programmes, the effect of major contextual factors, and the challenges of coordination among the different actors and the programmes they manage.

4.5 Evaluating sustainability and resilience in the SDGs

Sustainability concerns the ability of a community, programme or broader system to maintain equilibrium in the face of shocks and stresses. Resilience goes further by focusing on the ability of systems, entities, communities and individuals to withstand shocks and to recover quickly and effectively from catastrophes. Both sustainability and resilience involve assessing how a programme will respond to future events and making a judgment on how effectively it responded to past stresses and shocks.

Most of the proposed indicators for monitoring the progress of the SDGs, including those for gender and reducing inequalities, do not specifically measure sustainability or resilience. Sustainability evaluation means that different evaluation designs must be used that both continue to collect data over long periods of time and that have the flexibility to respond quickly and go back into the field to collect data

when shocks and crises occur. It may often be necessary to use naturalistic experiments that can take advantage of unanticipated shocks (floods, earthquakes, population displacements) to observe how groups or individuals cope.

All of these concerns also apply to EFGR evaluations as a central objective of both SDG-5 and SDG-10 is to strengthen sustainability and resilience. For example, SDG-5.2 seeks to eliminate all forms of violence against women in public and private spheres. As many gender interventions must combat deep-rooted beliefs and practices, interventions can have an initial success, which is then eroded by different forms of backlash. Consequently, it is essential to assess the extent to which improvements are sustained and that the community, and particularly women's organizations, have the resilience to learn from experience and promote the necessary changes to overcome earlier problems.

Promising approaches to evaluate equity-focused and gender-responsive dimensions of sustainability and resilience

The following are promising approaches to evaluate EFGR dimensions of sustainability and resilience:

a) Developing a theory of change that models sustainability and resilience: This would define the dimensions of sustainability and resilience relevant to gender and inequality and would model how different interventions would contribute to strengthening them.

b) Checklists to assess sustainability and resilience: There are a number of dimensions of sustainability and resilience that must be assessed. Depending on the nature of the programme, these may include: secure sources of finance for maintenance of infrastructure and the purchase of new equipment and supplies, regular procedures for inspection and maintenance, incentives for communities to continue to use services and to assist with maintenance, and continued political support.

c) Case studies: Both descriptive and QCA case studies are useful tools. QCA can assess the presence or absence of different sustainability and resilience dimensions and their effects on sustainability and overall programme performance. Descriptive case studies are useful for understanding the multiple social, economic and political factors that can undermine gender equality gains.

5. Proposed strategy to integrate equity-focused and gender-responsive evaluations into national SDG reviews

This chapter is based on the publication below. For additional information and details on how to integrate EFGR evaluations into national SDG reviews, we encourage you to consult it. The publication is available on the Internet at http://www.evalpartners.org/library/selected-books free of charge.

- Bamberger, Segone, and Reddy. 2014. "National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems". EvalPartners, UN Women and IOCE.

In addition, we invite you to take the related free e-learning programme available at: http://elearning.evalpartners.org/.

5.1 Integrating an equity-focused and gender-responsive lens into the national SDG evaluation policies and systems

While integrating EFGR lens into the national SDG evaluation policies and systems is very important, it can be quite challenging for the reasons described below. This chapter attempts to give indications on how to create a demand for this to happen.

Challenges

There are a number of challenges facing the design of the SDG evaluations with an EFGR lens. These can be classified into political and methodological.

Political challenges

When an EFGR-focused lens is used, many national development programmes will be rated less positively than when indices such as the Human Development Index, or other *aggregate measures* are used. Indices such as the Human Development Index estimate the average change in access to services (education, health, water supply) for the total population. There are many cases where the average accessibility score has improved but where certain groups have been left out. The national ratings appear much less favorable if they show (as is often the case) that the gap between the lowest economic and most vulnerable groups and the rest of the population has not decreased. In fact, as we saw in Chapter 3 (Section 3.1), there are many examples

where the equity gap has increased, even when average access has improved for the total population.

Many agencies are reluctant to accept this finding and consequently may resist this approach to evaluation. Also, EFGR evaluations often focus on socially marginalized groups (such as ethnic minorities, racially disadvantaged groups or other vulnerable groups such as Roma populations or refugees) that governments and society may not wish to support and may even be pressuring to leave (either areas where they live or even the country). A further political challenge is that in some countries it is not permitted to collect data by race, ethnicity or nationality.

Methodological challenges

EFGR evaluation requires disaggregation of data, which may be difficult. Also, some of the processes of social exclusion (such as being less welcoming to clients from certain ethnic or cultural groups) may be subtle and difficult to observe. This will sometimes require the use of participatory and qualitative techniques with which researchers may not be familiar. Social inclusion methodologies (see below) may also require the integration of different data sets (e.g., education, geographical location, income, ethnicity) which may be technically difficult. Many measures also require precise measurement of income and wealth, which is always difficult. Some measures also involve intrahousehold resource allocation, which is particularly difficult and expensive.

Reasons why many evaluations at the policy, programme and project levels do not address dimensions related to gender and to reducing inequalities

A principle of the evaluation of SDGs is that it is country-led and participatory, involving a wide range of government, civil society and community organizations as well as the national offices of international agencies. Consequently, it is important that the EFGR principles are fully integrated into the National Evaluation Policies (NEP) of each country. Bamberger et al. [66] estimated in 2014 that while only 16 developing countries had a formal NEP[67], many more regularly conduct gender-responsive evaluations, but not as part of an NEP. Of the 16 NEPs, only two referred directly to gender, but a number of other countries considered gender equality a development priority without it being included in the NEP. Similar data is not available on how many country evaluation programmes addressed equity-related issues.

66 Bamberger, M., Segone, M., and S. Reddy. 2014. "National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems". EvalPartners, UN Women and IOCE.

67 For a definition and discussion of national evaluation policies see Bamberger et al. 2014.

There are a number of reasons why gender equality is frequently not incorporated into NEPs (or into many other evaluations):

- Many NEPs only provide a high-level framework to ensure that government policies are systematically evaluated. These often focus on the process (e.g., when evaluations should be conducted) rather than the substance.
- Most NEPs do not include cross-cutting issues, such as gender equality, social equality or environment.
- Even in cases where gender equality issues are addressed in NEPs, the focus tends to be limited to monitoring women's access to services or their representation on community or political bodies, and there is usually little or no discussion of broader issues of power, access and participation, critical to advancing gender equality and women's human rights. Consequently, "women's issues" are considered very sector specific and only applicable in a narrow range of areas (such as women's health).
- In many cases, the NEP may not have existed at the time the gender policy was formulated and vice versa.

Box 8 also presents more general reasons why gender is frequently not incorporated into programme and project evaluations, all of which must be taken into consideration when developing strategies to ensure that SDG national evaluations are gender-responsive.

BOX 8. Reasons why gender equality issues are often not incorporated into policy, programme and project evaluations

- Some stakeholders do not see gender equality as relevant or useful. In some cases it is believed that a particular sector (such as roads, energy or trade) is "gender neutral" and that most women and men will benefit equally from the programmes—without the need for a gendered analysis in the evaluation.
- Gender evaluation methodologies or methodologies that are gender sensitive are unfamiliar to research and evaluation staff and they are perceived as being difficult to use.
- Additional costs and time may be required.
- Some stakeholders perceive gender as being threatening or likely to be unacceptable to certain stakeholders.
- Collecting gender-responsive data may be difficult or require extra work, potentially adding one more burden for overworked agency staff.
- Agencies with a strong quantitative background may perceive gender-analysis methods as being unprofessional or not rigorous.

Box 9 summarizes reasons why issues related to reducing inequalities are frequently not addressed in evaluations. Many of these are similar to the above discussion of gender (more complex methodologies, difficulties of access to data, and additional cost and time requirements). However, there are additional factors relating to the fact that a thorough examination of reducing inequalities issues (who is left behind?) tends to present a less positive image of national development than when only aggregate data on overall progress is presented.

So while large numbers of gender-responsive evaluations are conducted by many kinds of national and international agencies, it is clear that a strategy will be required in many countries to ensure that gender equality is fully integrated into the SDG national evaluation strategies. The same is true with respect to reducing inequalities.

BOX 9. Reasons why issues related to reducing inequalities are frequently not addressed in evaluations

- There is often political resistance to addressing equality issues or to providing additional benefits to marginal groups.
- Many countries are proud of the progress they are making to meet international development or to improve their ranking on the Human Development Index. When an EFGR lens is applied, it is often found that progress has been less satisfactory as the gap between the privileged and the vulnerable groups has not been reduced. Consequently, there may be political resistance to conducting or publishing these kinds of findings.
- Equity-focused evaluation is also technically more challenging, and some of the research methods are unfamiliar. They also require the use of more qualitative methods, which may be considered by some agencies to be less "professional" than the familiar quantitative methods.
- The costs of data collection and analysis may be higher and more time consuming.
- In addition, some of the required data, such as ethnicity, religion or physical and mental handicap, is difficult to collect as it is not included in conventional household income and expenditure surveys.
- Related to the previous point, many of the most vulnerable groups are largely invisible as they often do not have land or property titles or even identification cards. They often try to remain invisible for fear of reprisals from the police or other agencies.

Integrating an equity-focused and gender-responsive lens into the country-led SDG evaluation strategies

Box 10 presents lessons from the experience of international development agencies on the integration of a gender-responsive focus into evaluation. While these agencies have a different perspective from national governments, these lessons provide some useful pointers for the country-led evaluations. All of these lessons apply equally to the design and implementation of equity-focused evaluations.

BOX 10. Lessons from the experience of international agencies on the integration of gender into development evaluations

Lesson 1: Gender equality must be considered a central development objective in its own right, as well as essential to achieving other priorities and sectoral objectives.

Lesson 2: There is a need for a strong organizational commitment with direct senior management responsibility.

Lesson 3: There should be a clearly defined rationale for a gender-responsive development strategy. Different agencies prioritize different issues (for example, a business case or a human rights focus) and NEPs should clarify which rationale, or set of rationales, are prioritized.

Lesson 4: Guidelines, checklists and practical examples of gender-responsive evaluation are important.

Lesson 5: Alliance building and inter-agency coordination is also important.

Lesson 6: It should be recognized that gender interventions are frequently broad-based and complex, and that gender policies and evaluations must recognize this complexity.

Source: Bamberger et al. 2014, Chapter 3.

Flexibility is required to reflect the variations in how evaluation is organized in different countries. In some countries, there is a clearly defined NEP, while in others there may be active programmes of equity-focused or gender-responsive evaluations but no overall evaluation policy, and in still other countries, very few equity or gender focused evaluations are conducted. While recognizing the need for flexibility, the following are some general guidelines for incorporating gender and reduction of inequality into the country SDG evaluation strategies and NEP:

- Clearly define the rationale, objectives and purpose of the EFGR evaluations.
- Incorporate an EFGR lens into the guiding principles for the selection, conduct and use of evaluations.

- Seek consensus from stakeholders on the definition of EFGR evaluation, and a framework for their evaluation.

Box 5 (Chapter 3) gives examples of definitions of equity-focused evaluation and gender-responsive evaluation (taken from UNICEF and UN Women publications) that could be adapted and used as models. This should also involve agreement on the kinds of additional data and analysis that are required to adequately address these two themes.

Define responsibilities

Once the architecture of the SDG evaluation has been defined, it is important to ensure there is an explicit definition of how this applies to gender and reducing inequalities and who are the stakeholders and their responsibilities for these two areas. While in many countries, there are clearly defined responsibilities for gender, the same is not usually the case for reducing inequalities. Consequently, it will be important for each country to decide how reducing inequality should be addressed:

- Should there be a division within one particular agency (such as the Ministry of Planning or the Ministry of Social Development) that is responsible for incorporating an equity focus into the evaluations?
- Should each agency appoint a unit or an individual with this responsibility?
- Should each agency decide how to address reducing inequalities but with some overall guidelines?

Coverage and selection of evaluations

Evaluations of SDGs must address multiple issues, and there is a risk that gender and reducing inequalities may be ignored due to other more pressing priorities. Consequently, it is important to develop guidelines to ensure that a minimum set of issues related to gender and reducing inequalities are considered in most evaluations. If not possible, an alternative could be to put in place mechanisms to ensure that a few EFGR evaluations are included in the evaluation programme each year.

BOX 11. Developing gender-responsive evaluation guidelines in Colombia

The initiative to engender national M&E systems in the Latin America and Caribbean region started in 2012 with the project "Strengthening institutional capacities in monitoring and evaluation with a focus on gender equality, human rights and interculturality in Latin America and the Caribbean". The project started with a "Mapping of national M&E systems in Latin America and the Caribbean", which analysed the experience of institutionalizing evaluation in the region and the viability to integrate gender equality, human rights and interculturality dimensions into these systems. It followed with an international meeting of experts to present the findings of the mapping.

The meeting was attended by representatives from national M&E systems of Colombia (SINERGIA), Mexico (CONEVAL) and Peru (MIDIS, Ministry of Development and Social Inclusion) along with members of ReLAC (Latin American and the Caribbean Network for Monitoring, Evaluation and Systematization), CLEAR, Brazilian M&E Network, UNDP, United Nations Population Fund and United Nations Development Group – Latin America and the Caribbean. This step was instrumental for the creation of a Working Group on strengthening national evaluation capacities with a focus on gender equality and human rights as well as for the elaboration of a specific proposal for a systematic support to the national M&E systems and the regional and national evaluation networks and groups. This increased the demand for gender-responsive evaluation in the region, which resulted in the development of three training courses for staff of SINERGIA (Colombia), MIDIS (Peru) and CONEVAL (Mexico). In Colombia, SINERGIA went a step further by developing a full guidance document on integrating gender-responsive tools and approaches in their evaluation systems and processes.

The UN Women Regional Office for the Americas and the Caribbean and UN Women Country Office in Colombia are closely collaborating with SINERGIA to provide technical support in the integration of gender-responsive approaches into two evaluations of national public policies: the National Public Policy on Gender Equality (CONPES 161), and the National Policy on Prevention of Risks, Protection and Guarantee of Rights of Women Victims of the Armed Conflict (CONPES 3784). This is the first time UN Women is directly supporting a national evaluation system in mainstreaming a gender-responsive approach into the evaluation of major policies addressing gender equality and women's rights.

For more information, contact Laura Gonzalez (laura.gonzalez@unwomen.org).

Source: https://sinergia.dnp.gov.co/Paginas/Noticias/Sinergia_Capacitacion_ONU_Mujeres.aspx

Methodology

Existing methodological guidelines (like the one published by UNEG[68], UN Women[69] and UNICEF[70]) or new country-specific ones to be developed should be disseminated widely within countries to ensure that appropriate EFGR methodologies are used and that evaluation staff are trained in their use.

Ethics

While most agencies have guidelines for addressing ethical issues, it is important that these clearly address issues relating to gender, human rights and reducing inequalities:

- Ensure that measures are taken to understand and address religious, social, cultural and logistical contexts (e.g., access to transport and to mobile phones and internet) and, in some cases, political and administrative contexts that may constrain women and certain other groups (e.g., the disabled, ethnic minorities) from being interviewed or from speaking freely.
- Ensure that every effort is made to access data required for the evaluation to ensure that the situation of vulnerable groups can be studied and their situation compared to other mainstream groups. Be aware of strategies that can be used to avoid addressing these issues or making data available.[71]
- Attention should be paid to the emerging ethical challenges surrounding the access to and use of mobile phones and other new information technology. While information and communications technology make an important contribution to the empowerment of women and girls, there is a growing body of research on how, for example, mobile phones can increase domestic violence or cyber stalking.[72]
- New information technology also permits exciting new ways for women and young people (as well as low-income communities and other vulnerable groups) to provide anonymous feedback on

68 UN Evaluation Group. 2013. "Integrating human rights and gender equality in evaluation: Guidance document".

69 UN Women. 2015. "How to manage gender-responsive evaluation: Evaluation handbook". Independent Evaluation Office. New York, NY: UN Women.

70 Bamberger, M. and M. Segone. 2011. "How to design and manage equity focused evaluations". New York, NY: UNICEF.

71 For a discussion of ways agencies can avoid making data available, see Bamberger, Rugh and Mabry, Chapter 6.

72 Some of the potential negative consequences of increasing women's access to mobile phones include increased domestic violence (from jealous male partners who feel they are losing control), cyber-stalking and violent assaults against women who are seen carrying a mobile phone.

sensitive issues such as domestic violence, sexual harassment in schools, corruption and electoral fraud.[73]

- Require the use of ethical guidelines for research with the survivors of gender-based violence to try to protect women and girls from further discrimination or harm by participating in the evaluation process. Ensure that these guidelines also protect boys and men who are also subjected to gender-based violence.

Consultation and dissemination

Measures should be taken to ensure that all stakeholder groups and sectors of the population, specifically including women and vulnerable groups, have access to the evaluation studies and have the opportunity to participate, genuinely and systematically, in the consultation process. Mechanisms should be put in place to facilitate stakeholder feedback, including through mobile phones, internet and social media. It is also important to introduce creative (and context relevant and appropriate) forms of dissemination such as community meetings, dance, theatre and art for groups where conventional published reports are not appropriate. A critical, but often overlooked, issue is to ensure that reports are made available in local languages.

Use of the evaluation findings

Evaluation utilization continues to be one of the weakest areas of the evaluation process.[74] While it is essential to focus on the overall utilization of evaluation findings and recommendations, specific measures should be put in place to ensure gender-related findings, recommendations and lessons are also fully utilized. When promoting the dissemination of studies with an equity-focused and/or gender-responsive lens, it should be stressed that both of these are essential to ensure that "No-one is left behind".

5.2 Identifying stakeholders concerned with gender and reducing inequalities

The evaluation of SDGs will involve multiple stakeholders at the national, provincial and local levels as well as country offices of

73 For example, UNICEF has developed "U-REPORT" hand-held devices that have been used to allow school children to report on school-based issues such as trading sex for better exam results.

74 Bamberger, M. and K. MacKay. 2004 "Influential evaluations: Evaluations that improved performance and impacts of development programs". Operations Evaluation Department (now renamed Independent Evaluation Group). The World Bank. In 2005, a companion volume "Influential evaluations: Detailed case studies", was published.

international agencies. Within most of these groups, there are sub-groups or individuals who are specifically concerned with issues of gender and reducing inequality. It will be important to ensure that they are actively involved in the consultation on gender and reducing inequality, and that these specialists are also involved in the broader discussions of the SDG evaluations. There is sometimes a tendency to look upon gender as a specialized issue only of relevance to specialists and only relating to specific areas such as health and education but not consulted in areas such as infrastructure and trade.

5.3 Developing an equity-focused and gender-responsive national evaluation system

Below are some potential intervention points for developing an EFGR evaluation system.[75] For additional detailed information and an example of EFGR lens NEP, please consult Bamberger, Segone, and Reddy. 2014. "National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems". EvalPartners, UN Women and IOCE.

- Linking the evaluation policy to a national gender policy or strategy
- Highlighting international conventions on gender and women or social inclusion to which the country is a signatory
- Integrating an EFGR lens into the national development strategies
- Integrating an EFGR lens into poverty analysis
- Ensuring indicators for gender and for reducing inequalities are used in the national development strategy
- Incorporating an EFGR lens into social accountability strategies: Several NEPs include mechanisms for social accountability and citizen feedback. Sometimes this is achieved through a social observatory (e.g., Morocco) while in other countries, citizen participation is encouraged through mechanisms such as crowd-sourcing (e.g., Colombia) or citizen report cards.

75 For an example of a national, gender-focused evaluation system as it might be applied by a donor agency, see Bamberger, M. 2013. "Measuring and evaluating equity". Evaluation Conclave, Kathmandu. Workshop on Equity and Resilience.

5.4 Developing an advocacy strategy to promote equity-focused and gender-responsive evaluation use

In certain cases, an advocacy strategy is needed to ensure an EFGR lens is integrated into NEPs. Below are a few suggestions on how to developon how to develop and implement an advocacy campaign in this regard.

Typical steps in the design of an advocacy campaign

Karkara[76] identifies nine questions that must be addressed when developing an advocacy strategy for gender-responsive evaluation. The following list has been slightly adapted in order to apply to the promotion of both an EFGR evaluation strategy:

- What do we want to achieve? (goals)
- Who can give it to us? (audiences, power holders, opinion leaders)
- What do target audiences need to hear and what kinds of information will be convincing? (messages)
- Who do the target audiences need to hear it from? Who can deliver the message most effectively? (messengers)
- How do we get the target audiences to hear it? What is the most effective way to reach different audiences? (delivery)
- What have we got? (resources, strengths)
- What do we need to develop? (challenges, gaps)
- How do we begin? (first steps)
- How do we know if it is working? (M&E)

The questions and the precise focus vary depending on the focus of the campaign and the particular audience.

Figure 3 illustrates a strategy for identifying stakeholders with direct responsibility for policy formulation on EFGR evaluations, and stakeholders who can influence NEP policies and the types of influence that each group can have. It also identifies the types of influence that different kinds of stakeholders can have over the evaluation process. Following Mackay (2007), we propose three kinds of influence that stakeholders with direct influence can have: "carrots", which are positive incentives; "sticks", which are negative sanctions; and "sermons", which are different kinds of moral support. In contrast, stakeholders who do not have direct responsibility can influence the policies through financial and technical support, reference to international standards

76 Cited in Bamberger et al. 2014:64-68.

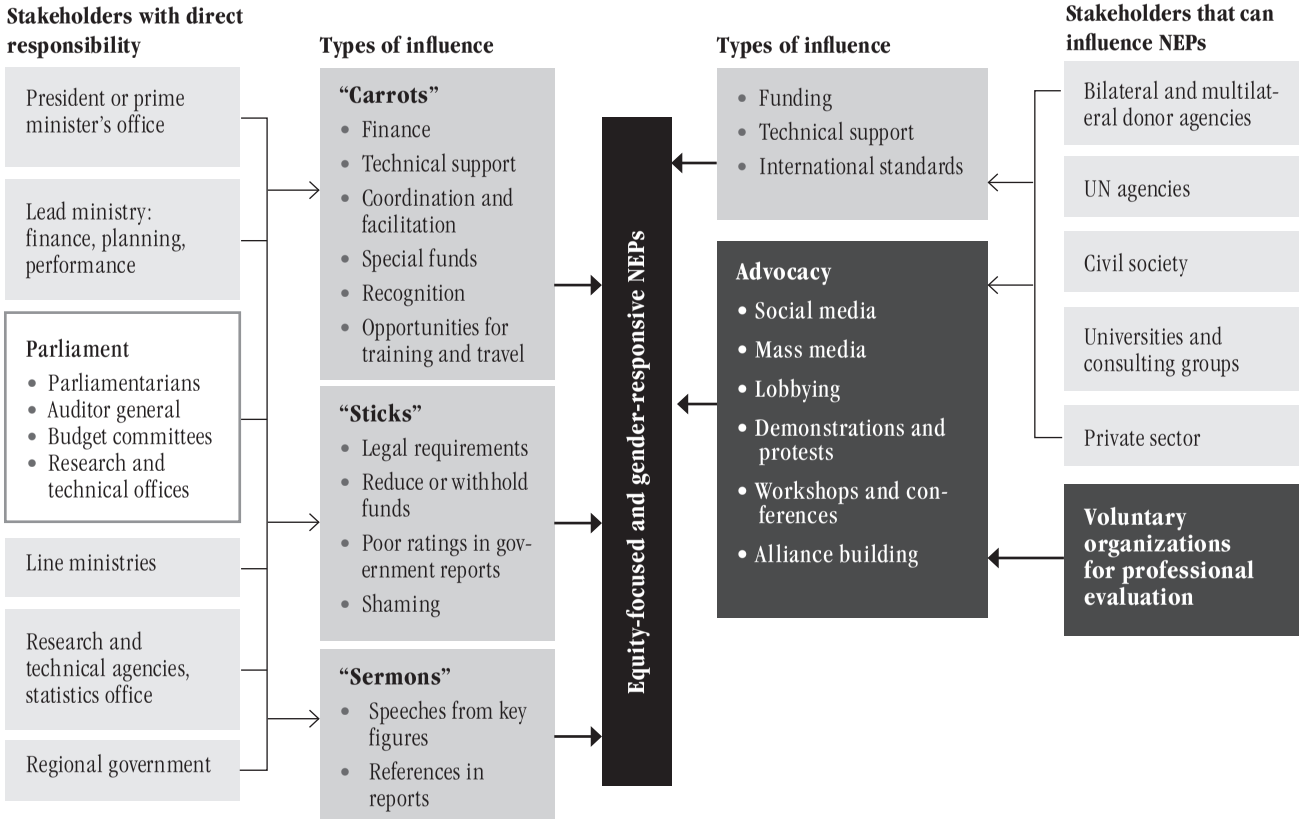
and different forms of advocacy. While all stakeholders use advocacy in one form or another, these are most widely used by civil society and voluntary organizations for professional evaluation.

BOX 12. An example of a campaign focused on parliamentarians: Potential entry points

Given that the SDG evaluations are voluntary and country driven, parliament is a key target for advocacy campaigns to promote EFGR. Parliaments have complex and unique decision-making processes, and it is important to identify key members, committees and staffers. The following are some of the potential entry points for a campaign:

- **Parliamentary leadership**: All leaders have their areas of personal interest, so it is important to identify who are the leaders most interested in issues related to gender and reduction of inequality, as well as use of evaluation for evidence-based policy making. It is also important to identify any key leaders who are opposed to these issues (for personal or political reasons).
- **Identifying key committees**: Most legislation is initiated by, or passes through, committees. The budget or finance committees also have a key role in most legislation.
- **Staff assistants**: Staff assistants often act as gatekeepers, and many issues and campaigns only reach the ears of parliamentarians through their staffers.
- **Caucuses**: These are groups of parliamentarians who share concerns on particular issues, geographical regions or population groups. They can be important allies.
- **Elections**: Elections provide one of the main opportunities to promote issues and to gain commitments from individual parliamentarians or groups. However, it is important to demonstrate that there are significant numbers of voters who will support candidates supporting these issues.
- **Pending legislation**: Advocacy groups need to learn well ahead of time when legislation is being proposed so that they can intervene at an early stage when ideas and options are still being explored.
- **Parliamentary research and documentation services**

Figure 3. Formulation of an equity-focused and gender-responsive evaluation advocacy strategy: Stakeholders with direct responsibility and stakeholders who can influence the process



Source: Bamberger, Segone, Reddy 2014, page 62.

Annex 1. Lessons from the MDGs

A1.1 The importance of learning lessons from the MDGs

While the scope has been significantly broadened, the SDGs are designed to continue and expand the initiatives launched by the MDGs. Consequently, there are many lessons to be learned from the MDGs with respect to the evaluation of the SDGs, and a number of agencies have conducted assessments of lessons learned from their experience with MDG evaluations. In this section, we summarize findings of the MDG assessment studies conducted by UNDP, the World Bank, the UN Office of Internal Oversight Service, the UN Development Group and the InterAgency and Expert Group Task Force. The present chapter highlights the key lessons of relevance to the SDG evaluation programme.

A1.2 Lessons for the overall framework of the SDG evaluations

i) It is important to develop a systematic and robust framework for monitoring and evaluation of the SDGs—this was lacking for the MDGs.

ii) More ambitious goals are required for middle-income countries. Many of these countries had already achieved many of the MDGs so the MDG goals became increasingly irrelevant for them.

iii) While some agencies believed the results framework proved an effective mechanism for identifying areas that were falling behind, others questioned the practical utility of results based management.[77] If a results framework is used, it is important to incorporate a results chain.

iv) More realistic and less ambitious SDG planning is needed. For example, more than half of UNDP's proposed MDG plans remained unfunded.

v) Many agencies found it difficult to integrate the MDGs into their own development strategies.

vi) It is important to focus on institutional strengthening at the country level.

77 The Office of Internal Oversight Services claimed that the results based management focus had little practical utility.

vii) There should be more focus on private sector development as this was largely ignored in the MDGs.

viii) The importance of a participatory development focus should also be recognized.

ix) There is a need for a clear, overarching M&E framework for the SDGs defining roles and responsibilities and coordinating mechanisms. This was lacking for the MDGs.

x) There is a need for a multi-tiered system linking country, regional and global M&E systems.

xi) The SDG framework should not only focus on UN agencies but also should include bilateral, multilateral and civil society agencies.

xii) There is a need to recognize and address differences among stakeholders with respect to accountability.

xiii) The new tools and techniques of the data revolution should be fully utilized.

xiv) One or more mechanisms, backed by considerable resources, will be needed to support statistical monitoring and, increasingly, evaluation capacity development.

xv) While it is important to nurture the distinct characteristics and strengths of different professional fields, it will be crucial to minimize the potential for overlap. This will require the coordination of outputs that feed into decision-making processes relating to the SDGs and the coordination of capacity-building, especially in the nascent area of evaluation.

xvi) Evaluation can be used as a bridge to: contextualize monitoring data, identify reasons why achievement appears to be thwarted, and increase the accessibility of evidence-based information for decision makers as well as other stakeholders.

A1.3 Pay more attention to equity, exclusion, gender equality and women's empowerment

i) Equity and social exclusion must play a much more central role than in the MDG evaluations.

ii) There is a need to go beyond the MDG focus on aggregate measures (proportion of the population with access to services) and to assess whether or not the gap between the poorest sectors and the rest of the population is closing.

iii) There is a need for a greater focus on participatory development.

iv) The targets and indicators for gender equality and women's empowerment need to be much broader than for the MDGs.

A1.4 Strengthen mechanisms for reporting, learning and advocacy

i) The MDG country scorecards were very specific and proved operationally useful. The SDGs should avoid the tendency of the MDG country evaluation reports to gradually become broader and less operationally useful.

ii) The MDG country reports played an important role in promoting debate on development issues so they should be given priority in the SDG evaluations.

iii) The MDGs found it was a challenge to keep the attention of the international community over such a long period of time. The SDGs should draw on some of the innovative dissemination techniques developed by a number of agencies.

iv) It will be a challenge to maintain the interest of all countries—not just the poorest.

v) The SDGs must plan for more systematic and sustained advocacy campaigns.

vi) More effective mechanisms to learn lessons and disseminate the findings must be developed.

vii) Monitoring data must be made more accessible.

A1.5 Evaluation capacity development

i) Three is a need to emphasize the importance of building statistical capacity in statistical agencies.

ii) Evaluation capacity development should not just target M&E offices but also agencies that commission, finance and use evaluations.

iii) Evaluation capacity development needs to strengthen qualitative evaluation as the MDGs tended to focus too narrowly on numerical outputs.

iv) A clear strategy is needed to define multilateral and bilateral support for evaluation capacity development.

A1.6 Coordination

i) Coordination of the many evaluation initiatives under the MDGs was weak and should be strengthened.

ii) Greater efforts are required to draw more effectively on specialized research and evaluation agencies at the country level.

iii) Central repositories must be created for evaluation findings.

A1.7 Methodology[78]

Summary of main points raised in the different assessments of the MDG evaluation methodology

i) There should be a greater focus on qualitative and mixed methods indicators and approaches, as the MDGs tend to only assess quantitative outcomes without regard to the quality of the services. In extreme cases, the reported services may either not be operating or are of such a poor quality as to be of little utility.

ii) There was a need for an integrated approach that examined the linkages among the different MDGs. There was a strong tendency towards a "silo approach" where each MDG was evaluated separately and with almost no reference to other MDGs, which could have a strong influence on outputs and outcomes.

iii) Organisation for Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC) criteria proved a useful way to assess progress and performance, although additional criteria will be required for the SDGs.

iv) The MDGs mainly focused on *outputs*, which are relatively easy to monitor and there was little focus on the more important and longer-term *outcomes*.

v) The evaluation would have benefited from the introduction of a theory of change that explained the processes through which interventions were expected to achieve their intended outputs and outcomes. This would have strengthened the ability of the evaluations to assess the extent to which programmes were successful.

78 The following lessons and recommendations are mainly drawn from the Inter-agency and Expert Task Force on "Lessons learned about MDG monitoring from a statistical perspective"; and United Nations. 2013. "Lessons learned from MDG monitoring from a statistical perspective".

vi) Although the term evaluation is frequently used in the MDGs, most of the focus was, in fact, on monitoring the delivery of the intended outputs.

vii) The focus on output measurement should have been complemented by process analysis to assess the effectiveness of the processes through which services were delivered. From an evaluation perspective, it is important to be able to assess when outputs or outcomes are not achieved, and whether or not this is due to *design failure* (weaknesses in the programme theory) or *implementation failure*.

Summary of points raised by the UN Task Force on MDG M&E

Weaknesses of the MDG framework from statistical and policy perspectives

i) Targets and indicators were perceived by national statistical systems and other development partners primarily as an international agency driven, "top-down" initiative.

ii) There were inconsistencies between goals, targets and indicators and a lack of clarity on how goals were set.

iii) Some of the numerical targets were too ambitious or poorly specified.

iv) Trends at global and regional levels tended to be dominated by a few countries with large populations.

v) The baseline year for many indicators was often set too far away from the framework's baseline year.

vi) The framework does not adequately address inequality issues between men and women, rural and urban areas, rich and poor, and among specific population groups.

vii) The statistical capacity in many countries is still limited. Data availability and data quality still remain a big challenge. In addition, the time lag between data collection and data dissemination is normally two to three years for most MDG indicators, which affects accountability and policymaking.

viii) Discrepancies between national and international data often created problems at the national level and tension in the international statistical community.

Criteria for setting targets

i) Greater clarity and a more systematic and consistent approach is required for target setting.

Criteria for indicator selection

A more systematic approach to indicator selection is required. The following criteria were proposed for the selection of indicators:

i) Relevance.

ii) Methodological soundness.

iii) Measurability: The indicator can be measured in a cost-effective and practical manner by all countries.

iv) Understandability: The indicator is clear and easy to understand for policymakers, the general public and other stakeholders.

Furthermore, the whole set of indicators should meet the following characteristics:

i) Coherence.

ii) Be limited in number.

iii) Have a greater focus on quality (and not just quantity).

Annex 2. Potential applications of big data and new information technologies for equity-focused and gender-responsive evaluations

The following table presents examples of the use of big data and new information technology with potential applications for the EFGR evaluations. Currently, most of the applications of these technologies are used for research, programme design and emergency relief rather than directly for programme evaluation—although many techniques could be adapted for programme evaluation. Some agencies are already using smart phones and similar technology for programme monitoring, but less frequently for programme evaluation. However, the techniques could easily be adapted for evaluation. For example, mapping studies or real-time early-warning data could easily be adapted to create baseline data for a pretest-posttest comparison group design.

Table 5. Big data and new information technology with potential applications for the EFGR evaluations

1. **Involving women in rural India in the design, implementation and dissemination of a tablet-based survey covering more than 800,000** households to identify community priorities. (Source: World Bank, India Social Observatory)[76]
2. **Identifying trends in discrimination against women in the workplace in Indonesia.** Online data was analysed to identify real-time signals of discrimination against women in the workplace in Indonesia. (Source: ILO in collaboration with the Indonesian Ministry of Development Planning [Bappenas] and UN Global Pulse)[77]
3. **Mapping poverty in China using call data records.** Call data records were mined to develop proxies for poverty indicators and could potentially provide a much more economical and continuous source of data on poverty trends. (Source: UNDP China Office 2014)
4. **The OCHA Ebola Map.** Office for the Coordination of Humanitarian Affairs integrated available data from different agencies and sources to provide regularly updated maps on the incidence and rates of spread of Ebola through West Africa, creating a standardized data platform consolidating data from different agencies. (Source: United Nations Office for the Coordination of Humanitarian Affairs)[78]

76 http://www.worldbank.org/en/programs/social-observatory

77 http://www.unglobalpulse.org/indonesia-women-employment

78 https://search.yahoo.com/search?fr=mcafee&type=C211US400D20150722&p=ocha+ebola+map

5. Using social media to map the Haitian earthquake in 2010 and to guide emergency services in locating victims. A group of volunteers analysed information from Facebook and Twitter, and later SMS messages relating to victims of the earthquake. The information was located on a rapidly constructed crisis street map that was posted to emergency relief agencies. More than 1.4 million edits were made to this map as information was refined. (Source: Patrick Meier "Digital Humanitarians")

6. Conducting radio-mining in Uganda. Data visualization was used to create maps permitting users to pinpoint locations where particular problems were identified through analysis of radio programmes. (Source UN Global Pulse)[79]

7. Using mobile phone data to track seasonal migration in Senegal. The movement of populations in Senegal was tracked using anonymized phone data. The results showed that for vulnerable populations, changes in mobility patterns could indicate changes in livelihoods or coping strategies. (Source: World Food Programme in collaboration with UN Global Pulse) [80]

8. Instant real-time data culled from social media to provide journalists with a crisis mapping platform with instant feedback on crises and emergencies, such as the 2007 elections in Kenya, and ongoing conflicts in Iraq and Syria, and Ebola in Liberia. (Source: Ushahidi) [81]

9. Evaluating the impacts of protected forest areas on forest cover in Mexico. A mixed method design was used to combine satellite images with on-the-ground data collection for a quasi-experimental design comparing protected forest areas where the agencies were working with other areas where they were not working. This design provided many more sources of data and improved the match through propensity score matching. The observations were compared at various points in time to assess differences in forest cover between the project and comparison areas. The mixed methods design permitted follow-up in-depth studies in Mexico, Brazil and other countries on the ground to dig deeper into the social, economic and political factors operating in different areas of illegal logging, illegal eco-tourism and other factors. (Source: Global Environmental Facility and UNDP)

79 http://unglobalpulse.org/radio-mining-uganda

80 http://www.unglobalpulse.org/projects/analysing-seasonal-mobility-patterns-using-mobile-phone-data

81 http://techpresident.com/news/wegov/25248/ushahidi-instant-crsis-data

Annex 3. Feedback from online consultations on the design of equity-focused and gender-responsive evaluations for the SDGs

In January and February 2016, EvalGender+ coordinated online consultations to provide guidance on the design of the EFGR component of the SDG evaluations. The following are some of the highlights of the consultations with respect to:

- Promising new metrics
- Useful resources on complexity-responsive evaluation

The full consultations in English, Spanish, Arabic and Russian are available at: http://gendereval.ning.com/forum/topics/eval-sdgs.

A3.1 Promising new metrics

General observations

- If we are looking at power relations, let us not forget those with other sexual identities like men who have sex with men and transgender.
- *Evaluation needs to be owned by the communities.* Extending beyond self-help groups and the government to focus on inclusively organized neighbourhood-groups could be a beginning for gender-responsive empowerment and governance participation. We should focus more on how communities can be key stakeholders while planning the evaluation and give them the ownership to decide on the indicators and tools. We can seek community perception on what has changed over 15 years (the SDG period) on gender and reduction of inequality issues: What has improved and what has not, for whom, and why?

What new metrics are more promising?

- Sustainability post-project evaluation needs to be done to measure sustainability, particularly from an equity-focused and gender equality lens.
- There is a need for facilitation skills in evaluation. While undertaking a study on reproductive health issues of migrant workers, asking them directly what are reproductive issues yielded discouraging

results. Therefore, the development of evaluator facilitation skills that can enhance stakeholder conversations and encourage deep listening are key.

- Participatory statistics through which communities can collect quantitative data:
  - Approaches to redefine and redistribute power are also central to potential for progress. Constellation's strength-based approach, SALT (Stimulate, Appreciate, Listen/Learn, Transfer), and self assessment framework is a way for communities to create a dream for themselves and also assess how far they have come on the path toward achieving that dream. It has been used in more than 60 countries and is being used by several agencies like the International Labour Organization (www.communitylifecompetence.org). Similarly, use of another strength-based approach, appreciative inquiry, can create an environment where stakeholders can share openly. A big challenge raised by the respondents was that during data collection, communities do not open and share on sensitive issues.
  - Listening: Are we really open to listen to what is coming from the informants? What kind of mind and heart should we have in order to listen and understand? Is it plausible to think that an open mind can determine an open discussion and no agenda, neither from the evaluator nor from the informants? Can an open mind lead to discussions that are closer to solutions or that are themselves solutions? We need a lot of courage, as human beings to accept that changes are happening in the present moment. It doesn't make any sense to reduce that moment to an "information gathering" process.
- Life course approach tools can help identify factors underlying inequalities.
- Systematization of experiences or strategic impact (CARE).

A 3.2 Useful resources for dealing with complexity

- ***Wicked solutions: A systems approach to complex problems.***
  Sample here: http://www.bobwilliams.co.nz/Systems_Resources
  html eBook version: http://www.gumroad.com/l/wicked
  Physical book version: http://bit.ly/1SVoOH3
- ***"Systems Concepts in Action" with Richard Hummelbrunner***
  Free systems, facilitation, organizational development and evaluation resources at: http://www.bobwilliams.co.nz

- CRS's ProPack II. Available at: http://www.crs.org/our-work-overseas/research-publications/propack-iii.
- *Dealing with Complexity in Development Evaluation, A Practical Approach* edited by Michael Bamberger, Jos Vaessen, and Estelle Raimondo.
- *Developmental Evaluation Exemplars: Principles in Practice*, edited by Michael Quinn Patton PhD, Kate McKegg and Nan Wehipeihana. Available at: http://www.amazon.com/Developmental-Evaluation-Exemplars-Principles-Practice/dp/1462522963/ref=mt_paperback?_encoding=UTF8&me=.
- *IDS Bulletin* devoted to complexity and systems thinking in evaluation: http://www.ids.ac.uk/publication/towards-systemic-approaches-to-evaluation-and-impact
- Sue Funnell and Patricia Rogers explored these issues, about how to respond to complexity, in the book on programme theory *Purposeful Programme Theory*.
- There's more detail on these in a conference presentation for Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)in 2011 and in a recent publication for the Office of the Chief Economist in Australia's Department of Industy, Innovation and Science "Choosing appropriate designs and methods for impact evaluation".
- Article from Nigel Simister, available at: http://www.intrac.org/data/files/resources/663/Developing-ME-Systems-for-Complex-Organisations.pdf.
- *Critical Systems Heuristics* by Reynolds & Williams, 2012, "Systems thinking and equity-focused evaluations".
- Burns, D. 2014. "Systemic action research: Changing systemic dynamics to support sustainable change". *Action Research Journal*, available at: http://arj.sagepub.com/content/12/1/3.full.pdf+html.
- "Navigating complexity in international development: Facilitating sustainable change at scale" by Danny Burns and Stuart Worsley, 2015.
- Personal experiences shared online at: http://gendereval.ning.com/forum/topics/eval-sdgs.

References

Bamberger, M. 2009. "Institutionalizing impact evaluation with the framework of a monitoring and evaluation system". Independent Evaluation Group. Washington, DC: World Bank.

Bamberger, M. 2013. "Measuring and evaluating equity". Evaluation Conclave, Kathmandu. Workshop on Equity and Resilience.

Bamberger, M. 2016. "Integrating big data into the monitoring and evaluation of development programmes". New York, NY: UN Global Pulse. (Scheduled for publication in October 2016)

Bamberger, M., and K. MacKay. 2004. "Influential evaluations: Evaluations that improved performance and impacts of development programs". Operations Evaluation Department (now renamed Independent Evaluation Group). Washington, DC: World Bank.

Bamberger, M., and K. MacKay. 2005. "Influential evaluations: Detailed case studies". Operations Evaluation Department. Washington, DC: World Bank.

Bamberger, M., Raftree, L., and V. Olazabal. 2016. "The role of new information and communication technologies in equity-focused evaluation: Opportunities and challenges". *Evaluation*, April 2016, 22 (2): 228-244.

Bamberger, M., Rugh, J., and L. Mabry. 2012. *Real World Evaluation: Working Under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage.

Bamberger, M. and M. Segone. 2011. "How to design and manage equity focused evaluations". New York, NY: UNICEF.

Bamberger, M., Segone, M., and S. Reddy. 2014. "National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems". EvalPartners, UN Women and IOCE. Available at: http://c.ymcdn.com/sites/www.afrea.org/resource/resmgr/books/NationalEvaluationPolicies_w.pdf.

Bamberger, M., Tarsilla, M., and S. Hesse-Biber. 2016. "Why so many 'rigorous' evaluations fail to identify unintended consequences of development programs. How mixed-methods can contribute". *Evaluation and Program Planning* 55: 155-162.

Bamberger, M., Vaessen, J., and E. Raimondo. 2016. *Dealing with Complexity in Development Evaluation: A Practical Approach*. Thousand Oaks, CA: Sage.

Byrne, D. and C. Ragin. 2009. *The Sage Handbook of Case-Based Methods*. Thousand Oaks, CA: Sage.

Byrne, D. 2009. "Complex realist and configurational approaches to cases: A radical synthesis." In Byrne and Ragin (op.cit)

Centre for the Study of Social Exclusion. http://sticerd.lse.ac.uk/case/

Cousins, J. and E. Whitmore. 1998. *Framing Participatory Evaluation: New Directions for Evaluation*. Wiley Periodicals.

Davies, R. and J. Dart. 2005. "The 'most significant change' (MSC) technique: A guide to its use". Melbourne. Available at: http://www.mande.co.uk/docs/MSCGuide.pdf.

Davis, A. and M. Guevara. 2016. "A case study in complexity: Evaluating a long-term effort to prevent gender-based violence in El Salvador". In Bamberger, Vaessen and Raimondo, editors. Op.cit.

Earle, S., Carden, F., and T. Smytlo. 2001. "Outcome mapping: Building learning and reflection into development programmes". Ottawa, Canada: International Development Research Center.

Fredman, S. and B. Goldblatt. 2015. "Gender equality and human rights". Discussion Paper. New York, NY: UN Women.

Funnell, S. and P. Rogers. 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass.

Hashemi, S., Schuler, S., and A. Riley. 1996. "Rural credit programmes and women's empowerment in Bangladesh." *World Development* 24(4): 635-53.

Hesse-Biber, S. No date. "Gender evaluation and feminist evaluation: What's the difference and why is this difference important to consider?" *AEA Feminist TIG Newsletter*.

Hesse-Biber, S., and M. Bamberger. 2013. Presentation in an AEA workshop on gender-responsive evaluation.

Kumar, S. 2002. "Methods for community participation: A complete guide for practitioners". London, UK: ITDG.

Kumar, S. 2013. "Understanding equity and growth from a development perspective: Background paper". New York, NY: Rockefeller Foundation.

Kurtz, C. F., and D.J. Snowden. 2003. "The new dynamics of strategy: Sense-making in a complex and complicated world". *IBM Systems Journal* 42(3): 462-483.

Krugman, J., Hanmer, L., Twigg, S., Hasan, T., McCleary-Sills, S., and J. Santamaria. 2014. "Voice and agency: Empowering women and girls for shared prosperity". Washington, DC: World Bank.

Kusek, J. and R. Rist. 2004. "Ten steps to a results-based monitoring and evaluation system". Washington, DC: World Bank.

Le Blanc, D. 2015. "Towards integration at last? The sustainable development goals as a network of targets". DESA working paper No. 141, March 2015.

Leeuw, F.L. 2016. "Linking theory-based evaluation and contribution analysis: Three problems and a few solutions". *Evaluatio*n, July 2012: 18 (3): 348-363.

MacKay, K. 2007. "How to build M&E systems to support better government". Independent Evaluation Group. Washington, DC: World Bank.

Mayne, J. 2011. "Contribution analysis: Addressing cause and effect". In Forss, K., Marra, M., and R. Schwartz. Eds., *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers.

Mertens, D. M., and S. Hesse-Biber. 2013. "Mixed methods and credibility of evidence in evaluation". *New Directions for Evaluation*, 2013: 5–13. doi: 10.1002/ev.20053.

Overholt, C., Anderson, M., Cloud, K., and J. Austin. 1985. *Gender Roles in Development Projects: A Case Book*. Connecticut: Kumarian.

Patton, M. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, NY: Guilford Press.

Rogers, P. 2011. "Implications of complicated and complex characteristics for key tasks in evaluation". In Forss, K., Marra, M., and R. Schwartz. Eds., *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers.

Salmen, L. 1987. *Listen to the People: Participant Observer Evaluation of Development Programs*. New York, NY: Oxford University Press.

Segone, M., Bamberger, M., and U. Nagel. 2013. "How to design and manage equity focused evaluations". Presentation at the South Asia Evaluation Conclave. Kathmandu, November 2013.

Shadish, W., Cook, T., and D. Campbell. *Experimental and Quasi-experimental Designs for Field Settings*. Boston, MA: Houghton Mifflin.

Siegel, E. 2013. *Predictive Analytics: The Power to Predict who will Click, Buy, Lie or Die*. New Jersey: Wiley.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and B. Befani. 2012. "Broadening the range of designs and methods for impact evaluation". Working Paper No 38. London, UK: Department of International Development.

United Nations. 2015. "Transforming our world: The 2030 Agenda for Sustainable Development".

UN Development Group. 2015. "Guidance note: On supporting the tailoring of the 2030 Agenda for Sustainable Development to nation contexts". Version 1.0.

UN Development Programme. 2015. "Towards a baseline study: Insights on national evaluation capacities in 43 countries". Independent Evaluation Office. New York, NY: UNDP.

UN Development Programme. 2014. "Big data for development in China". Available at: http://www.cn.undp.org/content/china/en/home/library/south-south-cooperation/big-data-for-development-in-china.html

UN Economic and Social Council. 2015. "Report of the Inter-agency Expert Group on sustainable development goal indicators".

UN Environmental Programme. Undated. "Universality in the post 2015 Sustainable Development Agenda". Post 2015 Note #9 / OHCHR Human Rights and Post 2015. pp. 1-2 (undated). Available at: http://www.ohchr.org/Documents/Issues/MDGs/Post2015/OHCHR_UNEP.pdf.

UN Evaluation Group. 2013. "Integrating human rights and gender equality in evaluation: Guidance document".

UN Secretary-General. 2016. "Report of the Secretary-General: Critical milestones towards coherent, efficient and inclusive follow-up and review at the global level".

UN Women. 2014. "An empowered future: Corporate evaluation of UN Women's contribution to women's economic empowerment". Independent Evaluation Office. New York, NY: UN Women.

UN Women. 2015a. "Monitoring gender equality and the empowerment of women and girls in the 2030 Agenda for Sustainable Development: Opportunities and challenges". Position paper, August 2015. New York, NY: UN Women.

UN Women. 2015b. "Monitoring gender equality and the empowerment of women in the 2030 Agenda for Sustainable Development: Opportunities and challenges. A position paper". September 2015. New York, NY: UN Women.

UN Women. 2015c. "How to manage gender-responsive evaluation: Evaluation handbook". Independent Evaluation Office. New York, NY: UN Women.

USAID. 2013. "Complexity-aware monitoring". Washington, DC: USAID.

Valadez, J., and M. Bamberger. 1994. "Monitoring and evaluating social programs in developing countries". Washington, DC: World Bank.

Westley, F., Zimmerman B., and M.Q. Patton. 2006. "Getting to maybe: How the world is changed". Toronto, Ontario: Random House Canada.

White, H. 2009. "Theory-based impact evaluation: Principles and practice". *Journal of Development Effectiveness*, 1 (3): 271-284.

Williams, B., and R. Hummelbrunner. 2011. *Systems Concepts in Action: A Practitioner's Toolkit*. Palo Alto, CA: Stanford University Press.

Williams, B. 2011. "All methods are wrong, some methods are useful". *Systems Thinker* 22(4).

Wilson-Grau, R., and H. Britt. 2012. *Outcome Harvesting*. Cairo, Egypt: The Ford Foundation's Middle East and North Africa Office. Available at: http://www.outcomemapping.ca/resource/resource.php?id=374.

World Bank. 2012. "Gender equality and development". World Development Report. Washington, DC: World Bank.

Yin, R. 2012. *Applications of Case Study Research*. Thousand Oaks, CA: Sage.

Zolli, A. 2012. *Resilience: Why Things Bounce Back*. New York, NY: Free Press.

The new 2030 Agenda for Sustainable Development shapes the sustainable and equitable development discourse and action for the next 15 years. The 2030 Agenda calls for follow-up and review mechanisms to ensure the Sustainable Development Goals are systematically monitored and reviewed to ensure "No one is left behind." Evaluation plays a crucial role to support effective and efficient implementation of the Sustainable Development Goals as the 2030 Agenda reviews are to be informed by country-led evaluations and joint evaluations. Evaluation capacity-development support is key for countries to strengthen national data and evaluation systems. Evaluation is an important source of evidence on how policies, national strategies and programmes delivered results and what needs to be done differently. But is this enough to ensure no one is left behind? This practical book provides guidance and tools to contribute to evaluation practice that is transformative and responsive to inclusiveness, participation and ownership. It presents guidance for applying an equity-focused and gender-responsive approach to evaluation and help make certain the Sustainable Development Goals are evaluated to ensure "No one is left behind".

2016