GENDER-SENSITIVE EVALUATION

BEST AND PROMISING PRACTICES IN ENGENDERING EVALUATION

SEPTEMBER 2014

Toolkit for Integrating LGBT Activities into Programming in the E&E Region September 2014, Doc ID: PA-00K-43Z

Gender-sensitive Evaluation: Best and Promising Practices in Engendering Evaluation September 2014, Doc ID: PA-00K-43P

Equal Rights, Equal Justice: Toolkit for Addressing Gender-Based Violence Through Rule of Law Projects September 2014, Doc ID: PA-00K-439

Case Management Toolkit: A User's Guide for Strengthening Case Management Services in Child Welfare July 2014, Doc ID: PA-00J-ZSZ

TVET Models, Structures and Policy Reform: Evidence from the Europe & Eurasia Region July 2014, Doc ID: PA-00J-ZSW

Grounding the Advocacy Argument: Measuring the National Budget Effects of Improving Roma Education Levels in Albania, Serbia and Republic of Macedonia May 2014, Doc ID: PA-00J-ZT4

Study on Deinstitutionalization of Children and Adults with Disabilities in Europe and Eurasia December 2013, Doc ID: PA-00K-2R5

Testing the Waters: LGBT People in the Europe & Eurasia Region November 2013, Doc ID: PN-AEC-993

Gender Analysis of Agriculture in the Europe and Eurasia Region November 2013, Doc ID: PA-00K-2R6

Promising Approaches – Addressing Child Trafficking in Europe and Eurasia February 2013, Doc ID: PN-AEB-806

Domestic Violence in Europe and Eurasia—2012 Update August 2012, Doc ID: PA-00H-QCC

Women with Disabilities in the Europe & Eurasia Region August 2012, Doc ID: PN-ADZ-326

Toward Gender Equality in Europe and Eurasia: A Toolkit for Analysis May 2012, Doc ID: PN-AEA-292

Tackling the Demand that Fosters Human Trafficking August 2011, Doc ID: PN-ADZ-753

The Other Side of the Gender Equation: Gender Issues for Men in the Europe and Eurasia Region July 2011, Doc ID: PN-ADZ-757

Education Vulnerability Analysis for the E&E Region September 2010, Doc ID: PN-ADX-519

Trafficking of Adult Men in the Europe and Eurasia Region June 2010, Doc ID: PN-ADW-368

Out of the Broken Mirror: Learning for Reconciliation through Multi-perspective History Teaching in Southeast Europe May 2010, Doc ID: PD-ACS-531

Best Practices in Inclusive Education for Children with Disabilities: Application for Program Design in the Europe and Eurasia Region March 2010, Doc ID: PA-00H-PH4

An Evaluation Framework for USAID-Funded TIP Prevention and Victim Protection Programs December 2009, Doc ID: PN-ADR-430

Transitions towards an Inclusive Future: Vocational Skills Development and Employment Options for Persons with Disabilities in Europe & Eurasia October 2009, Doc ID: PN-ADS-499

The Prevalence of Disability in Europe & Eurasia September 2009, Doc ID: PN-ADT-697

Toolkit for Integrating Domestic Violence Activities into Programming in Europe and Eurasia July 2009, Doc ID: PN-ADO-921

The Job that Remains: An Overview of USAID Child Welfare Reform Efforts in Europe and Eurasia June 2009, Doc ID: PN-ADO-922

Best Practices in Trafficking Prevention in Europe and Eurasia January 2009, Doc ID: PN-ADO-543, PN-ADO-765

Methods and Models for Mixing Services for Victims of Domestic Violence and Trafficking in Persons in Europe and Eurasia December 2008, Doc ID: PN-ADN-297

Social Work Education and the Practice Environment in Europe and Eurasia December 2008, Doc ID: PN-ADN-298

Best Practices for Programming to Protect and Assist Victims of Trafficking in Europe and Eurasia November 2008, Doc ID: PN-ADN-295

GENDER-SENSITIVE EVALUATION AND MONITORING

BEST AND PROMISING PRACTICES IN ENGENDERING EVALUATIONS

Prepared by

Sharon Brisolara for JBS International, Inc.

September 2014

TABLE OF CONTENTS

TABLE OF FIGURES

EXECUTIVE SUMMARY

USAID's Gender Equality and Female Empowerment Policy is a clear statement of the Agency's commitment to "improve the lives of citizens around the world by advancing equality between females and males, and empowering women and girls to participate fully in and benefit from the development of their societies" (USAID, 2012). Achieving gender equality and female empowerment are Millennium Development Goals (MDGs), and both are vehicles for achieving several of the other MDGs. International development agencies engaged in work to reduce persistent gender gaps have learned a great deal over the past decade about the concerted efforts required and the importance of attending to gender throughout the life of a project.

To engender an evaluation is to view the evaluand (that which is being evaluated) through the lens of gender; it is to ensure that gender needs, realities, and issues are consistently and specifically being considered at each stage of the evaluation. Gender-sensitive evaluations are responsive to the fact that gender disparities are shaped and reinforced by cultural values, norms, structures, institutions, policies, organizational practices, and programs. By using gender-sensitive approaches, evaluation can be a key tool for understanding the structural causes of gender inequalities and the consequences of such inequalities for women and men.

Quality program evaluations are critical to ensuring the effectiveness, efficiency, and sustainability of development efforts and the achievement of gender equality goals. Engendering an evaluation often requires specific attention to the language and concepts that are sometimes implicit in a project design. Gender equality aims can be compromised in the evaluation design or at various stages of its implementation. Often, evaluators discover insufficient attention to gender indicators or data gaps that inhibit the ability to set baselines, measure progress, or understand dynamics and unintended outcomes. Gender-sensitive evaluations account for the possibility of such barriers; engage gender, cultural, and methodological experts; and recognize the need to foster leadership and accountability in mainstreaming attention to gender equality.

The work of international aid organizations and evaluation and research on gender equality efforts have led to a range of best and promising practices with respect to successfully engendering evaluations. In preparing to conduct a program evaluation, these good practices include integrating a gender analysis into an evaluability assessment, selecting or creating clear gender-sensitive indicators, adopting mixed-method evaluation designs, thoroughly reviewing existing M&E systems, and accessing other reliable sources of pertinent gender equality data. Another important step in preparing for an engendered evaluation is ensuring that the Scope of Work clarifies the gender equality focus of the program (and sector- and country-level gender equality goals, where applicable) and includes specific language regarding the composition and responsibilities of the evaluation team.

Gender-sensitive training has been found to be important not only for the evaluation team, but for staff and the range of stakeholders invested in the success of the project. Clarifying concepts and terms, explaining the purpose of engendering evaluation, and ensuring equality of representation should all be addressed in such training. Indeed, bringing gender awareness to considerations of representation and inclusion is a hallmark of gender-sensitive evaluation. Salient demographic and identifying characteristics will differ according to cultural context and sector; however, the development of procedures to ensure representation and participation is critical in considering the formation of the evaluation team and identifying evaluation respondents.

Once the purpose of the evaluation is determined, selection of an appropriate evaluation model should reflect the intended use(s) of the evaluation results and who will be taking action based on these findings. Models most suitable for engendered evaluations value the involvement of stakeholders to

some degree in the evaluation and consider project participants to be important project stakeholders. The model(s) chosen guide the framing of the key evaluation questions, questions which focus on differential impacts for various actors by gender as the result of the program. Within gender-sensitive evaluations, project theories of change reflect the structural, systemic nature of gender inequalities, the complex nature of power dynamics, and the non-linear nature of change. Throughout the evaluation, such issues require that attention is paid to unintended and unanticipated consequences, particularly for the most vulnerable populations, organized by gender.

Gender-sensitive data collection is not so much about specific instruments, but about the sensitivity of instruments and evaluators to the situation and conditions of respondents. Data collection protocols and the items within instruments must also be gender-sensitive and seek to unearth the nature of existing gender gaps and differential gender impacts of USAID programs. Annex A provides checklists of engendered questions to assist in the engendering process, and diverse perspectives should be brought to bear on the analysis of data and the interpretation of findings. Annex A offers a guide for the evaluation team and a structure with which to apply, throughout the evaluation, a gendered lens.

Gender-Sensitive Evaluation: Best and Promising Practices in Engendering Evaluations

"[T]o improve the lives of citizens around the world by advancing equality between females and males, and empowering women and girls to participate fully in and benefit from the development of their societies…through integration of gender equality and female empowerment throughout the Agency's Program Cycle and related processes: in strategic planning, programming, project design and implementation, and monitoring and evaluation."

USAID Gender Equality and Female Empowerment Policy, 2012

I. INTRODUCTION

Much has been written on the critical role of gender equality and/or female empowerment in creating sustainable development results even prior to the establishment of the United Nations Millennium Development Goals (MDGs) in 2000. Donor agencies working in the field of international development have adopted policies and practices that reflect the realization that gender equality is central to long-term, effective development in all sectors, and that promoting gender equality and empowering women (MDG 3) is a vehicle for achieving many of the other MDGs, including universal primary education, improving maternal health, combating HIV/AIDS, and decreasing child mortality rates (World Bank, 2007). In establishing its *Gender Equality and Female Empowerment Policy*, USAID reaffirmed its commitment to reducing gender gaps in critical development outcomes and contributing to the empowerment of women and girls. In addition to this policy, USAID has invested in gender equality assessment tools and the production of this document in order to create internal guides and "How To" documents on engendering USAID evaluations.

Sustainability, effectiveness, significance of impact - these are practical concerns. Equality objectives are also based on values, values of fairness and social justice, of the importance of human rights, and of individual wellbeing (Segone & Bamberger, 2013). In some cases, programs and policies work to foster equality by focusing on the needs of the most disadvantaged groups as well as ensuring that the benefits of the program, or of participation, are equally available to all participants regardless of the various dimensions of their identities (Segone & Bamberger, 2013).

While there are multiple determinants of inequalities, and many such determinants intersect with each other, gender remains a key dimension. Recent decades have seen the construction of multiple frameworks and gender analysis techniques for understanding and assessing the gendered dimensions of development goals. And yet, despite this increased focus on gender, gender disparities persist. Attempts to integrate gender equality goals throughout policies and programs in international development efforts have often been found to be inconsistent or partial (African Development Bank, 2011). Research on effective ways of mainstreaming or integrating gender equality and female empowerment goals into development work have emphasized the importance of these goals being reflected in institutional policies, hiring practices, country-level plans, the planning of initiatives, program design, budgeting, and monitoring and evaluation systems and practices.

A country's monitoring and evaluation (M&E) system plays a significant role in achieving a commitment to gender equality and female empowerment. A well-designed M&E system assesses the extent to which gender-equitable development goals have been achieved and promotes consistent attention to a range of development goals. Indeed, fostering the development and promotion of gender-sensitive evaluations has been identified as one of five key priorities of Eval Partners in 2012 (Kosheleva & Segone, 2013). Eval Partners, self-described as a global movement to strengthen national evaluation capacities,

views this priority as part of an overall commitment to equality, human rights, and social justice. International donor agencies, for similar reasons, have been active in exploring effective ways of integrating gender equality efforts throughout their missions. These efforts have resulted in a body of research, evaluations, and reflections that offer practitioners a range of best and promising practices related to fostering gender equality and female empowerment.

The purpose of this document is to consolidate best practices for USAID staff, contractors, and stakeholders on how to prepare for and implement evaluations that reflect a sensitivity to gender concerns, examine the extent to which interventions produce positive, equitable results for people regardless of gender, and assess the extent to which addressing gender gaps or issues leads to better development results.

II. GENDER-SENSITIVE EVALUATION: ENGENDERING EVALUATION

To engender an evaluation is to view the evaluation process, dynamics, design, and the key elements of each evaluation stage through a "gender lens" to make sure that the evaluation and associated data collection and analysis practices are fully informed by an awareness of how gender shapes and is shaped by both programs and evaluations. By engendering an evaluation, all aspects of that evaluation deeply consider gender needs and issues; the evaluation is responsive to the particular cultural context in which gendered relationships play out; and inquiry is focused on the data and analysis needed to achieve gender equality and empowerment objectives. Understanding the concept of engendering evaluation requires clarity regarding three key concepts: what constitutes gender, what is meant by an evaluation that is gender-sensitive, and what role evaluation plays in achieving gender equality.

GENDER

As indicated by USAID's policy on gender equality, gender is not synonymous with a focus on women. The policy is explicitly "inclusive of all women and men, girls and boys, regardless of age, sexual orientation, gender identity, disability status, religion, ethnicity, socioeconomic status, geographic area, migratory status, forced displacement, or HIV/AIDS status" (USAID, 2012, p.2). To speak of gender is to refer to socially constructed and contextualized roles and responsibilities that are separate from biology or an individual's sex. Women, men, boys and girls are socialized in particular ways specific to their cultures; they are shaped by different social norms and are affected in different ways by policies, interventions, interactions with others, and socio-political environments (CARE, 2012).

GENDER-SENSITIVE

Gender-sensitive evaluation is an approach to evaluation that pays specific and sustained attention to gender needs, interests, and culturally specific dynamics and recognizes the disparities in opportunities, resources, and power that are organized by gender and that are pervasive. Further, an evaluation that is gender-sensitive is responsive to the fact that not only are gender and gender disparities shaped and reinforced by cultural values and norms, but also by structures, institutions, policies, organizational practices, and programs. As such, gender-sensitive evaluation has become a key tool for understanding the structural causes of gender inequalities and the consequences of such inequalities for men and women (Espinosa, 2013; De Waal, 2006). What is the nature of gender relations? What gender inequalities exist and what are the structural causes of inequality, particularly within the program context? What are the gender effects of structural conditions, policies, and programs? What can be done to reduce gender inequalities and empower both women and men to more fully engage in social

and political life? (Espinosa, 2013, 2011). These are questions that gender-sensitive evaluations strive to answer.

Gender-sensitive evaluation can also be explained in contrast with gender-blind strategies. Gender-blind strategies, whether they be evaluations or policies, do not consider the effect of gender, gender norms, differential power, or unequal access to resources on participants or, importantly, on potential program outcomes. Such strategies also do not account for how program activities or outcomes can affect, positively or negatively, gender dynamics and conditions. Gender-sensitive evaluations begin with having explored gender relations and explicitly recognize and attend to gender and gender equality concerns throughout the evaluation process.

A gender-sensitive approach to evaluation is one in which all stages of the evaluation reflect: (a) an awareness that the degree and meaning of program participation, program results, and potential sustainability are shaped by gender; (b) a recognition that explicit attention to gender issues must be integrated into the evaluation if gender equality objectives are to be addressed; and (c) a commitment to examining the extent to which gender equality was achieved as a result of the program (intervention). Fully integrating a gender-sensitive approach means that these elements are present in the scope of work (SOW) used in soliciting an evaluation, in steps leading to the evaluation design, in the choice of methods, data collection tools and protocols, and throughout data analysis and interpretation. Furthermore, attention to these criteria should be reflected in the choice of reporting strategy as well as in the utilization plan that results from the evaluation findings. To attend faithfully and specifically to gender in this way is to engender an evaluation.

Often, a program focused on gender equality or female empowerment will attend explicitly to gender issues or disparities. The program goals, the outputs, outcomes, and impacts, the target population, and program activities are in service of such equality. Staff members are trained in the theory and reasons underlying this work. Program activities typically directly address gender concerns and are clearly based on a gender analysis. However, a gender-sensitive approach can be more difficult when it comes to programs and initiatives that do not have an explicit gender equality or female empowerment focus. While agency policies make clear that attention to gender equality should extend beyond individual programs, sometimes such attention is only reflected in a superficial way, such as the requirement that certain data be disaggregated or in the articulation of target participation rates by gender. If a Mission is committed to achieving greater gender equality through its collective efforts, gender-sensitive program design, informed by and assessed by gender-sensitive evaluation practices, is of critical importance for all evaluations. The recommendations in this document reflect best and promising practices for engendering evaluations and conducting gender-sensitive evaluations regardless of whether or not gender equality and female empowerment are the central objective of the program. Given USAID's policies, practices and commitments, engendering evaluations is not an approach reserved for specific programs; it is good evaluation practice.[1]

THE ROLE OF EVALUATION

Let us be specific about the role of evaluation and monitoring systems. While program evaluation is concerned with accountability to donor agencies, governments, and stakeholders, its value extends beyond that. Program evaluation contributes to the effectiveness and sustainability of efforts by highlighting positive results and the activities or conditions that led to these results. Evaluation also

---

1. It may not always be possible to follow all recommendations for every evaluation given budget or timing realities; there may be cases, for example, when hiring a gender equality expert is beyond the scope of the budget given the country location. In such cases, it is important to prioritize which evaluations are most in need of a closer gender-sensitive examination and to see how the evaluation can otherwise address gender-sensitive evaluation concerns (e.g., seeing in what other ways gender expertise can be reflected in the evaluation).

contributes to unearthing negative consequences so that these can be avoided or mitigated, determines capacity-building needs, and assesses local sustainability and commitment needs within a particular community context. Practitioners working with an engendered evaluation approach as an integral component of program designs have noted that a commitment to engendering evaluation contributes to the creation of baseline data collection, appropriate project-specific monitoring indicators, clear evaluation procedures throughout the project cycle, and increased evaluation capacity (Lexow & Hansen, 2005). Well-designed country-led evaluation systems and functions are vital to ensuring that development interventions are effective, efficient, responsive, and improve the quality of life for all participants regardless of gender (Espinosa, 2013; Kosheleva & Segone, 2013; World Bank, 2005).

Nevertheless, program and M&E staff members often encounter barriers to taking advantage of the full range of benefits evaluations. Designing evaluations in concert with program planning strengthens program design and the program's ability to achieve results; however, it is not uncommon to encounter an evaluation solicitation that is produced after the program has been designed or implemented. How one ensures that evaluations are gender-sensitive differs with the type of program and focus of the evaluation as well. Evaluations of gender equality or female empowerment programs explicitly focus on gender outcomes because those are the program's main goals. For other types of programs, however, gender equality may not be a direct outcome and the evaluation questions may not explicitly refer to gender. Even in these cases, however, gender needs to be considered within the scope of the evaluation and the evaluation team needs to address gender inequalities in relation to the questions asked to ensure that future programming takes into account any unanticipated outcomes or emerging gender disparities. Gender-sensitive evaluation approaches and guidance materials must be poised to address the evaluation needs of both types of programs.

The particular qualities and criteria for what constitutes a "good" evaluation will vary according to international development context and donor agency. Typically, however, evaluation criteria include attention to rigor (concerns about validity, reliability, adequacy to the task at hand, addressing attribution), efficiency (designs that are economical, timely, simple, targeted), measurability (by quantitative and qualitative means, can be monitored), credibility (to a range of stakeholders), and clarity (perceived as clear, specific, understandable) (Morra, Imas, & Rist, 2009; World Bank, 2008). Increasingly, practitioners have noted that credible evaluations must demonstrate critical awareness of the cultural context (Kirkhart, 2010) and have designs flexible enough to respond to shifting realities while remaining focused on the intended use by intended users (Bamberger, Rugh, & Mabry, 2012; Patton, 2011). Furthermore, development practitioners working with equality-focused efforts have recommended an assessment of the impact and sustainability of interventions with specific reference to equitable development results (Bamberger & Segone, 2011).

The next section of this document addresses the barriers that can inhibit the engendering of an evaluation in order to bring a greater awareness to the issues that must be addressed by a gender-sensitive evaluation. The final section of the document presents best and promising practices in creating gender-sensitive evaluation approaches.

III. KEY CHALLENGES AND CONSIDERATIONS IN ENGENDERING EVALUATIONS

The literature on designing engendered evaluations reveals common challenges faced by program staff and evaluation practitioners that have specific evaluation consequences. These can be broadly categorized as issues related to Language and Concepts, Program and Project Design, Monitoring and Evaluation Practices, Competencies and Expertise, Ethical Considerations, and Leadership and Accountability. This section highlights the key elements of these challenges as a foundation from which

stakeholders can better understand how to respond to and design evaluations in order to overcome such challenges.

LANGUAGE AND CONCEPT CHALLENGES

In an earlier section, terms such as gender and gender-sensitive were defined. In designing a gender-sensitive evaluation, it is important that a range of key concepts are clarified and that they are clarified within the context of the country and program of interest. Without such attention to key terms and concepts, program planners, evaluation practitioners, and other stakeholders can operate from different assumptions resulting in, at best, omissions, confusion, and inefficiencies. Without conceptual clarity, it is difficult to fully take into account existing gendered norms (division of labor is one example); such clarity is needed in order to avoid reinforcing gender inequalities or causing unintentional harm (CARE, 2012).

While it is true that Agency policies specifically define all key terms and that these appear in documents available to staff and stakeholders, program staff, management, evaluation participants, and others can become overwhelmed by being faced with multiple policies and procedures. Furthermore, evaluation team members may not have been trained in or exposed to all relevant policies, or may not remember or have focused specifically on relevant policies. Fortunately, there are strategies for addressing this issue as the section on best and promising practices attests.

The following key terms and concepts may be helpful in considering how evaluations can be engendered.

- Gender is an inclusive term and refers to women, men, girls and boys and their relation to each other.

- Gender refers to socially constructed roles, taboos, prohibitions and restrictions, expectations, available occupations, division of labor and responsibilities, and opportunities available interpersonally, within the household, community and at regional and national levels. Evaluations should frame findings within the gender relations in the particular local context.

- Gender interacts with other salient dimensions of identity that can also lead to marginalization; these include class, caste, race, ethnicity, age, disability status, sexuality, sexual orientation, gender identity, religion, socioeconomic status, geographic area of residence, migratory or citizenship status, forced displacement, or HIV/AIDS status (Bustelo, 2010; CARE, 2012; Sielbeck-Bowen et al, 2002; USAID, 2012). Power is not distributed equally among these different dimensions (FAO, 2011); different subgroups might require different data collection strategies, protocols, or accommodations.

- Gender equality, at its core, is concerned with the idea that rights and opportunities do not depend on one being male or female (CIDA, 2010). Specific gender equality goals should be clearly articulated at the country, sector, and program levels, including larger interests such as shifts in attitudes or transformed power relations between and among women and men (Ruxton, 2004).

- In cases where female empowerment is a desired outcome, clarity about what empowerment means and does not mean, the goals of empowerment, and the strategies that are being used to achieve those ends is critical (UN HABITAT, 2012). It is important that the understandings (as there may be more than one) of such concepts are made explicit so that evaluation questions and outcomes can be appropriately focused and readers can appropriately assess the significance of findings.

- Gender analysis is a term that can be used to describe a range of strategies or activities. At one level, it is a systematic attempt to identify key issues contributing to gender inequalities and how

gendered power relations lead to discrimination, subordination and exclusion (Vainio-Mattila 2010; CIDA, 2010). USAID policy discusses the domains of analysis to include in gender analyses (USAID, 2013). It is important that the evaluation team knows what constitutes a gender analysis, whether or not a gender analysis was conducted prior to program implementation, what specific tools were/can be used to conduct a gender analysis for the program in question, and what was learned through the analysis.

PROJECT DESIGN

Even when standards state that all projects are to be designed with gender equality in mind, or that a gender analysis should be done and reflected in the logical framework matrix (or logic model), evaluators may find that this is not consistently the case. If gender equality goals have not been sufficiently integrated into the program design, there can be significant implications for the evaluation's timeline and budget. The following are key issues that might be encountered:

- *Design issues*: Project activities may not address key gender issues, may not be based on an analysis of local needs, or may not explicitly reflect gender-equality objectives.

- *Implementation issues*: Gender equality goals may not have been fully implemented or may have been selectively implemented due to staffing issues, delays in start-up, budget shortfalls, or lack of resources. Similarly, the program may not have been implemented with fidelity to its design.

- *Skills/expertise*: Evaluation teams may not include or utilize individuals with the required skills or expertise at the time of the evaluation. Best practices for engendering evaluation involve including gender experts, culturally competent practitioners, local language skills, methodological experts, and leaders credible to participants.

- *Intensity*: The project design may reflect gender-sensitive approaches, but activities may not have been sufficiently long or frequent enough to affect desired changes.

- *Recruitment/Participation*: Gender equality goals may fall apart as the result of issues with recruitment or ability to participate: segments of the population may be missing due to ineffective or incomplete recruitment efforts or barriers to accessing the program (such as the indirect costs of participation, transportation issues, and stigma).

MONITORING AND EVALUATION PRACTICES

Gender-sensitive indicators, sex-disaggregated data, and attention to gender relations are required elements of USAID evaluations (USAID, 2011); researchers and practitioners have affirmed the need for attention to all of these elements in fostering a gender-sensitive evaluation. Experience in the field has shown that despite clear policies, evaluation practitioners often find disparities between the monitoring data being collected at the Mission level and the needs of individual programs, particularly when evaluators did not participate in the initial program design. If collected data cannot be disaggregated by sex, gender equality outcomes or claims can be difficult, if not impossible, to assess. Using proxy data may not shed sufficient light on the effectiveness of activities, presenting a challenge to the evaluation's ability to make strong or useful claims. Particular and common M&E issues can negatively affect the extent to which evaluations can be engendered. The list below highlights key issues.

- *Insufficient attention to appropriate, local gender indicators and issues.* This situation can result from a narrow or unsystematic gender analysis at the project or sector level. It is important to ask if both the analysis and evaluation plan are inclusive of the range of factors that affect gender. These include social, economic, political, legal, institutional, and psychological/interpersonal factors (Bamberger, 2013) at the country level that will manifest in different ways for individual

sectors. Is gender reflected in outputs, outcomes, and impacts?

- *Insufficient attention to the diversity of perspectives and participants.* It is important to examine, within the M&E system as well as in the individual evaluation, which perspectives and experiences are missing (Gender, Agriculture, & Assets Project, 2012). It is also important to remember that women and men are not, even within the same country, homogenous entities (CIDA, 2012).

- *Insufficient attention to unintended outcomes.* Even when intended outcomes are adequately tracked, gender-equality goals can be compromised if an assessment of unintended outcomes is not also included within the evaluation (Bamberger, 2013). An awareness of unintended negative consequences of development projects for women has contributed to the present need for gender equality evaluations, yet M&E systems and evaluations often overlook this important information need.

- *Limited capacity to implement the M&E plan.* Implementation capacity can hinder the best of plans or make it difficult to access data when needed. Insufficient staff, inadequate training in key concepts and systems, inadequate dedicated budget, and difficulty accessing data (especially in a timely way) due to incomplete documentation or storage situations are all possible issues. While engendered evaluations will likely incur initial additional costs related to data disaggregation, the tracking of additional indicators, and stakeholder training, it is important to remember that: a) there are also social and economic costs of not integrating gender equality goals into projects and evaluations; and b) gender equality and female empowerment involve complex changes that occur over time (UNEG, 2011).

- *Issues related to rigor and credibility.* Examinations of M&E systems often discuss the need for rigorous quantitative data and baseline data. And yet, within gender-sensitive evaluations, qualitative data is also needed to assess many difficult to measure issues (Bamberger, 2013). In addition, recognizing gender equality results often calls for historical and other data suitable for drawing comparisons or adequately addressing attribution (OSCE, 2011). This might include data on potential participants who have not yet enrolled in or benefited from the program, data from comparable programs, or access to national population based data and indices that examine gender inequalities (Paluck & Ball, 2010). Similarly, it is important to ensure that any issues related to the integrity of indicators, particularly gender-related indicators, are addressed. What constitutes a particular indicator may have changed over time, compromising the ability to make comparisons.

COMPETENCIES AND EXPERTISE

While all evaluations draw on cultural competencies and methodological expertise, gender-sensitive evaluations have particular needs that require specific areas of expertise. Needed expertise can be broadly categorized as being related to gender in general as well as methodological and evaluation skills and cultural competencies related to the gender issues being examined. These skills become more important given the constraints that evaluators typically face with respect to time, data, budget, and politicized situations (Bamberger, Rugh, & Mabry, 2012) and are especially important to the collection of gender-sensitive data.

- *Gender expertise*: Not having a gender expert available at the design of the evaluation (and/or program) can result in the omission of important gender-related questions and indicators. Gender expertise is also very useful in analyzing and interpreting data, in framing the search for unintended or negative consequences for individuals and in communities. Even when a program and its corresponding evaluation is rooted in a thorough gender analysis, it is important to have

a staff person with a dedicated focus on gender issues to examine evaluation plans, methods, and data with an engendered lens.

- *Cultural competency*: Gender socialization and resulting gendered roles and responsibilities are culturally and historically contingent. Said another way, gender and gender roles, responsibilities, dynamics, and opportunities must be understood within their cultural context. Without a deep knowledge of cultural norms, values, and socialization processes, evaluations risk misinterpreting data, ignoring or not collecting data on salient gender-related issues, excluding important groups or imposing other/Western values and priorities, or even contributing to negative or harmful gender consequences.

- *Methodological and evaluation expertise*: Engendered evaluations typically include use of both quantitative and qualitative methods. The evaluation team should include expertise in creating participatory processes by including diverse stakeholders in various stages of the evaluation's development, and in selecting data collection tools and protocols appropriate to gender norms, cultural values, and educational/literacy levels of respondents.

ETHICAL CONSIDERATIONS

Ethical considerations are not unique to gender-sensitive evaluations; rather, gender-specific questions and perspectives expand the realm of ethical considerations. (For a rich description of principles of good evaluation practice, see the American Evaluation Association's Guiding Principles for Evaluators, http://www.eval. org/p/cm/ld/fid=51.) The key ethical issues in gender-sensitive evaluations have to do with unintended consequences, inclusion, and transparency.

Unintended Consequences. The importance of examining and being conscious of unintended consequences on gender roles, identity, and empowerment was mentioned above. Unanticipated outcomes might be positive shifts in attitudes and dynamics or harmful, negative effects. It is important to note that such consequences might arise in the short-term or after the conclusion of the intervention. In addition to making space for examining unintended consequences, it is important to consider how negative consequences will be communicated and addressed.

Inclusion. Best and promising practices indicate the importance of participation in evaluation processes and, as noted above, of diverse segments of stakeholders. To speak of inclusion also raises the issue of how the evaluation team members will negotiate among differing versions or perspectives of empowerment, desired objectives, gender equality, cultural norms, or who should be included among stakeholders and respondents. The AEA's guiding principles state that "Where feasible, evaluators should attempt to foster social equity in evaluation, so that those who give to the evaluation may benefit in return." Making sure that diverse and typically underrepresented populations are included in the evaluation, ideally as both representatives and participants in evaluation activities, is therefore an important ethical issue.

Transparency. Addressing gender inequalities often requires attention to complex and sensitive issues such as gendered violence, genital mutilation, and laws related to inheritance, divorce, marriage, and education that negatively affect women or men. Evaluations can fail to address important elements of transparency related to communication of general and specific objectives (including the gender-sensitive nature of the evaluation), culturally unacceptable or negative results or consequences, and implications or potential uses of the findings. In addition to posing ethical risks, failing to address these elements risks undermining the credibility and utility of the evaluation (World Bank, 2008).

LEADERSHIP AND ACCOUNTABILITY

A final impediment to a fully engendered evaluation is insufficient knowledge and interest among key management and program staff. Without a sound understanding of the importance of gender equality and female empowerment goals, of gender-sensitive processes, of the engendered nature of programs, projects, policies, and institutions, and a concomitant commitment in attending to these, it is difficult to create and follow through on an engendered evaluation (World Bank, 2008). Given the multiple policy requirements for which managers and program staff are responsible, it can be difficult to develop such a commitment without explicit accountability frameworks that outline gender equality goals. Such accountability language is not always present in evaluation documents.

IV. BEST AND PROMISING PRACTICES AND CONSIDERATIONS

Given the importance and visibility of gender equality as a key development goal, donor agencies, researchers, and practitioners have adopted many new practices in order to engender evaluations. Not all of these practices have been subsequently evaluated, but most draw on rich experiences and observations in diverse contexts; some are based on meta-evaluations and extended studies.

Engendering evaluations requires attention at every stage of the evaluation. Practices related to each of these stages have been addressed separately as a means of highlighting key issues arising at each stage. These best and promising practices have been organized as follows:

Evaluation Preparation

- Determining the Evaluability of the Evaluation
- Selecting Clear Objectives and Indicators
- Reviewing Monitoring and Evaluation Systems
- Preparing an Engendered Scope of Work
- Presenting the Need to Engender an Evaluation to Staff and Stakeholders
- Selecting the Evaluation Team

Evaluation Design and Implementation

- Evaluation Questions
- Evaluation Model
- Theory of Change
- Data Collection: Tools, Protocols, and Methods
- Analysis and Interpretation
- Reporting and Utilization

EVALUATION PREPARATION

The steps leading up to an evaluation are critically important and set the foundation for the extent to which an evaluation will be gender-sensitive. For the purposes of this document, we begin our examination of evaluation preparedness at the point at which an evaluation of a program, project, or initiative is deemed necessary. The section on key challenges and considerations provides the context for many of these recommendations as well as issues important to address at the Operating Unit (OU) level in order to enhance evaluation capacity in general.

DETERMINING THE EVALUABILITY OF A PROGRAM AND CONDUCTING A GENDER ANALYSIS

An evaluability assessment is an exercise that helps to identify whether an intervention can be evaluated and whether an evaluation is justified, feasible, and likely to provide useful information. Its purpose is not only to conclude whether or not the evaluation can be undertaken, but also to guide further preparation of the intervention so that all of the necessary conditions for evaluation are in place (UNEG, 2011). Ideally, this assessment occurs prior to contracting with an evaluator in order to determine the needs of the evaluation or preparation stages; some evaluation practitioners might build such an assessment and responses into early stages of the evaluation.

Within program evaluation, evaluability assessments are typically qualitative efforts that examine whether or not:

- Program goals (including gender equality goals), objectives, and priority information needs are well defined and agreed upon;

- Program goals and objectives are plausible and realistic given resources and implementation specifics;

- Relevant performance data are available or can be made available for a reasonable cost;

- Intended users of the evaluation results are identified and agree on how the evaluation results will be used (Rossi, Lipsey, & Freeman, 2004).

A gender analysis is important in determining evaluability if the evaluation is to be gender-sensitive. Gender analysis is one of two mandatory analyses, the results of which are to be integrated throughout planning, design, and evaluation stages of projects (USAID, 2012a). USAID's Gender Equality and Female Empowerment Policy describes gender analysis as follows (p. 12):

Gender analysis is a tool for examining the differences between the roles that women and men play in communities and societies, the different levels of power they hold, their differing needs, constraints and opportunities, and the impact of these differences on their lives. . . At the strategy and project level, the gender analysis should identify root causes of existing gender inequalities or obstacles to female empowerment in that context . . . The gender analysis should also identify potential adverse impacts and/or risks of gender based exclusion that could result from planned activities.

While many gender analysis frameworks exist, each with its own strengths and weaknesses (see Warren, 2007 for an overview; Lexow & Hansen, 2005), conducting a gender analysis does not have to imply a high degree of complexity. It does, however, require posing key questions, and searching for quantitative and or qualitative data/information that will answer these questions. (See ADS 205 for specific questions and strategies.)

Reviewing the gender analysis completed for the project and assessing the quality of that analysis (e.g., have the most relevant aspects of a gender analysis been included?) is an important step in preparing for evaluation. If a gender analysis has not been conducted, how will the evaluation respond to this situation? Potential responses include:

- Conduct a literature review of relevant, culturally appropriate, research and gender analyses focusing on projects in similar sectors implemented by USAID or other donors.

- Focus on the missing elements of a gender analysis related to gender roles and responsibilities. Elements include: (1) laws, policies, regulations, and institutional practices; (2) cultural norms and beliefs; (3) gender roles, responsibilities, and time use; (4) access to and control over assets and resources; and (5) patterns of power and decision-making.

- Ensure the full participation of gender experts rooted in the given cultural context during all stages of the evaluation.

- Gather qualitative data through interviews, focus groups, or similar methods from representatives of the diverse populations served by the program.

SELECTING CLEAR OBJECTIVES AND INDICATORS

Selecting key gender-sensitive indicators arises in the literature as a matter of concern for those undertaking engendered evaluations. An indicator is a quantitative or qualitative variable that provides a simple measure of inputs, outputs, or outcomes used to define desired targets to be achieved, to assess progress at a given time, to measure change, and, ultimately, to evaluate the extent to which project results have been accomplished. Gender-sensitive indicators further assess the degree to which these results have been equitable and have contributed to gender equality goals. M&E systems typically consist of regular and systematic collection of indicators related to inputs, outputs, and short-term and long-term outcomes.

Inputs refer to resources invested in the delivery of a program or project.

Outputs generally refer to the products, activities, participation, or services that result from project activities.

Outcomes are desired, specific, realistic, and measureable program results. These constitute the most important indicators because they reflect the extent to which a particular project goal was achieved.

Evaluations and reviews related to the creation of gender-sensitive indicators have resulted in the following recommendations:

- Focus on a few key indicators with an emphasis on outcome indicators.

- Link indicators directly to the program's objectives, desired outcomes, and key activities (see ADS 205). Given finite resources, indicators should focus on the most critical issues indicated by the gender analysis that can be realistically measured within the project's timeframe and budget. Disaggregating data, training in data collection, and the reviews of indicators do result in costs that will need to be reflected in M&E budgets.

- Review existing indicators related to the sector/subject matter of interest and the program's regional and cultural context. Select those most relevant to the program's context (see ADS 205.3.6.2). Much work has been done by donor agencies related to sector specific gender indicators.

- Develop new indicators specific to program results in a participatory process to ensure that these are both relevant to the gender realities in the given setting and understandable to intended users (see ADS 205.3.6.1). Participation should ideally include gender experts, program staff, evaluation team members, community members, program participants, and other key stakeholders (CIDA, 1997; NDF, 2004; UN HABITAT, 2012).

- Ensure that both quantitative and qualitative indicators are included and that indicators provide information necessary to examining gender gaps, challenges, exclusions, strengths, and opportunities for greater gender equality and female empowerment (see ADS 205). Quantitative indicators by themselves are insufficient to capture complex issues such as women's empowerment or participation (Reinharz, 1992). More recently, development practitioners have urged a greater focus on perceptions of change among stakeholders using qualitative indicators (Kusek & Rist, 2004).

- Account for the diversity that exists among women and men locally when creating indicators. In addition to being disaggregated by sex, indicators should be able to be disaggregated by age, ethnicity, ability, socio-economic status, and other variables important to the agency, the program, and communities.

- Make sure that, within M&E systems, gender-sensitive indicators measure gender-related changes in communities and societies over time (CIDA, 1997).

- Check that indicators are technically sound, clearly defined, easy to use, and easy to understand (see ADS 203).

Donor agencies have adopted different criteria or critical tests over time for making sure that indicators or broader objectives are technically sound and clear. A common acronym used to describe these criteria is SMART. That is, objectives should be specific, measureable, achievable, reliable, and time bound. A more recent acronym is CREAM; indicators should be clear (unambiguous), relevant (to the subject and context at hand), economic (for a reasonable cost), adequate (sufficient to conduct the assessment), and monitorable (can be measured and validated) (Kusek & Rist, 2004). USAID requires that data and the processes use to collect them ensure the validity, integrity, precision, reliability, and timeliness of the data. (For more information, see ADS 203.3.11.1.) Gender-sensitive evaluations are well-advised to attend to such criteria. The United Nations and the Organization for Economic Cooperation and Development's Development Assistance Committee have added specific gender equality dimensions to their evaluation criteria, including equality, participation, social transformation, inclusiveness, and empowerment (UNEG, 2011).

REVIEWING MONITORING AND EVALUATION SYSTEMS

Finally, M&E practitioners must ensure that they gather baseline data to describe problems or conditions prior to a given intervention, an issue that demonstrates the importance of understanding and attending to key gender equality indicators from the onset of the project. Finally, in order to integrate gender equality throughout M&E and organizational systems, it is important that all staff, including senior management, recognizes that gender equality is the responsibility of everyone (Monbart & Rathner, 2010; NIBR, 2006). Indeed, gender equality requires an environment that supports such efforts; ideally, such an environment includes incentives, tracking, and accountability frameworks (UN AIDS, 2012).

A key point made in examinations of evaluation systems is that, in addition to considering what indicators need to be added and in what form, it is important to note that data collection frameworks and instruments themselves often reflect a gender bias (Corner, 2003) and thus can obscure key gender issues. A thorough review is useful in searching for such issues as well as in reinforcing the idea that gender equality goals should be reflected throughout M&E systems, not only in cases of gender equality or female empowerment projects.

A thorough review of data collection tools and frameworks using a gender-sensitive perspective or lens is another important step in making these systems gender-sensitive. Using a gendered lens will be discussed in the section on evaluation design as well as in the accompanying checklists. One element of this review should consist of reviewing coding and classification systems and terminologies. Do codes and terms reflect what was discovered in the gender analysis? Are additional codes or categories required? Do terms exclude or ignore groups or subgroups?

M&E systems would benefit from specific gender training for those involved in data collection and data entry, for supervisors as well as those actually responsible for coding data (Corner, 2003). Such training can result in another point of data quality control, in minimizing the possibility that data entry staff fix what they assume to be errors, and so that missing data can be identified. Furthermore, practitioners

recommend having both women and men involved in data collection, coding, and establishing engendered M&E systems.

In discussing indicators and rigorous evaluation, development and evaluation practitioners often point to the need for comparison data. As previously noted, evaluations may be able to gather data from comparison programs or may have access to participants signed up for, but not yet receiving an intervention. Practical reasons and ethical concerns can make an experimental or even a quasi-experimental design difficult to achieve. However, M&E systems can make use of other sources of gender equality data when local comparison data is unavailable. [The first 5 bullets below were compiled by Bamberger (2013).] Examples of these databases include:

- *UNDP Human Development Index and the Gender Inequality Index*: Includes information on gender differences in life expectancy, access to health, education and labor markets, and political participation. http://hdr.undp.org/en/statistics/gii

- *The UNDP Multiple Dimension Poverty Index*: Comparative data available for male and female-headed households. http://hdr.undp.org/en/statistics/mpi

- *MDG databases*: Can provide cross-country comparative, sex-disaggregated data on topics such as access to health, education, water supply, and sanitation. http://mdgs.un.org/unsd/mdg/Default.aspx

- *Demographic and Health Surveys (DHS), UNICEF's Multiple Indicator Cluster Survey (MICS), the World Bank Living Standards Measurement Study (LSMS)*: Comparative cross-country databases that provide sex-disaggregated socioeconomic data at the national, regional, and, often, community level. For the MICS, go to http://www.unicef.org/statistics/index_24302.html; for the LSMS, http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,content MDK:21610833~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html.

- *The Organization for Economic Co-operation and Development's (OECD) Social Institutions* and *Gender Index*: Countries are given high, low, or medium scores on gender equality indices. http://www.oecd.org/dev/poverty/theoecdsocialinstitutionsandgenderindex.htm

- *National Public Health databases or regional databases such as the African Gender and Development Index* can also be useful sources of disaggregated data. In some cases, national evaluation organizations, identified through the International Organization for Cooperation in Evaluation (IOCE), may also include local level public health or socio-economic data sources. For more on the African Gender and Development Index, see http://www.uneca.org/sites/default/files/publications/agdi_book_final.pdf.

PREPARING AN ENGENDERED EVALUATION SCOPE OF WORK

The scope of work (SOW) for a specific evaluation is a critical element of establishing evaluation expectations including the gender-sensitive framework for the evaluation. The more clarity that can be provided in the SOW, request for proposal, and similar documents, the more focused and useful the evaluation is likely to be.

The following list describes some of the recommended strategies for creating an engendered SOW. [Note that issues related to competency and expertise are covered earlier in this document.]

- Define the scope and purpose of the evaluation, including the particular gender equality focus and gender-sensitive indicators (Bamberger & Segone, 2011; Lexow & Hansen, 2005).

- Clearly identify the range of evaluation stakeholders and the range of project participants (the respondents, including those who are hard to reach; identify all subgroups, attending to gender

within each subgroup). Clarify who should be included in the evaluation (as respondents) and the composition (in terms of positions and competencies) of the evaluation team.

- Clarify the uses to which the evaluation results will be put. In other words, what are the intended uses and who are the intended users? How do these uses differ by gender within each group? If different users require specific report formats, this should be accounted for in the deliverables and reflected within the budget.

- Specify the need for a mixed method evaluation, if appropriate, that includes quantitative and qualitative data to enhance the ability to triangulate data sources, enhance the credibility of the evaluation, and to examine complex and changing gender realities.

- Identify the resources that the evaluation team will have at its disposal to complete a gender-sensitive evaluation. Specify and provide access to any gender analysis, gender action plan, theory of change description, and/or stakeholder analyses that were completed. Clearly identify key cultural issues and the potential and specific challenges that the team might encounter in implementing a gender equality-focused evaluation within a given context, project, or organization (Bamberger & Segone, 2011).

PRESENTING THE NEED TO ENGENDER EVALUATION

Ensuring that stakeholders and all staff, not just those involved in the evaluation, share a common understanding of what it means to engender an evaluation is integral to creating a suitable environment for an engendered evaluation. Fostering a common understanding serves many purposes. Without explicit discussions, misunderstandings can occur; even more harmful is a situation in which staff members assume that they are making decisions based on the same definitions when they may not be. With clarity about the elements and reasons for engendering evaluation, it is more likely that stakeholders, including participants, will disseminate similar information and explanations about the evaluation's objectives and will use evaluation results. Increasing the number of participants cognizant of the goals of gender equality, particularly at the level of the evaluation, enhances the chances that gender-blind or harmful elements will be recognized during the planning stage, and that critical cultural and gender issues will be reflected in the design.

In designing a training on engendering evaluations, consider the following:

- Avoid being perceived as "the gender police" (ILO, 2007). Emphasize that the training and engendered evaluation both contribute to a deeper understand of the organization's capacity for promoting gender equality. Ensure that the training helps identify where additional attention is needed.

- Clearly define all key terms using easily understood language. While some staff may be gender or M&E experts themselves, the training will likely include new staff, stakeholders new to such discussions, and individuals who use different definitions of these terms. Some of the terms that will be important to define include gender, gender equality, empowerment, evaluation, gender-sensitive, and engendered.

- Remember to provide a context for the presentation. Explain, for example, evaluations will examine the outcomes for both men and women to ensure that programming achieves the best results for everyone. If possible, use examples from local programs or examples from other locations with similar social norms (Espeln, 2006; Kanesathasan, et al, 2013).

- When possible, involve both men and women in facilitating these discussions and presentations. Having presenters share their personal commitment to gender equality and asking participants, perhaps at the end, about their own commitments or thoughts is also useful. Allow for meaningful participation in discussions about evaluation materials. (CARE, 2012).

SELECTING THE EVALUATION TEAM

The composition of the evaluation team has significant repercussions for the focus of the evaluation. The discussion of this composition has been divided into issues of expertise and representation.

As suggested earlier, there are three primary areas of expertise that should be reflected in the SOW and in the evaluation team:

Gender expertise: The team should include at least one gender expert. Organizational/Mission standards may specify what constitutes a gender expert. At the very least, the gender expert should have had experience in designing or leading engendered evaluations or programs and in conducting gender analysis or feminist evaluation, and should be able to demonstrate an understanding of gender issues, gender roles, and gender responsibilities (OECD, 2010).

Cultural expertise: While it is advisable to ensure that all evaluation team members demonstrate cultural competence (see the American Evaluation Association's statement on cultural Competence at http://www.eval.org/p/cm/ld/fid=92), it is also important to have specific cultural expertise represented on the evaluation team. Cultural experts should have strong knowledge about the local culture, including an awareness of norms, subgroups and interactions among groups, cultural practices and perspectives, and key gender issues. Cultural experts will likely be instrumental in identifying respondents and others who should be represented in the evaluation and language needs.

Methodological expertise: Given the recommendations that engendered evaluations implement mixed-method designs, it is important that the evaluation team demonstrate expertise in quantitative and qualitative methods. To engage key stakeholders in the evaluation in meaningful ways, it is helpful to include evaluators who demonstrate experience with and appreciation for participatory approaches and methods. Obviously, having a lead member with program evaluation experience (as opposed to research experience) is very important in designing a focused, engendered evaluation.

In addition to the areas of expertise mentioned above, it is important to consider issues of representation in the composition of the evaluation team. In terms of engendering evaluations, it is recommended that the team include both women and men, at least one person with knowledge or experience of the sector (or content area), and that a majority of the team demonstrates a commitment to gender equality. While OUs may have to make certain trade-offs as a result of resource constraints, this last recommendation is important for ensuring the evaluation is engendered throughout the implementation of the evaluation, data analysis, and report writing.

EVALUATION DESIGN AND IMPLEMENTATION

The preceding section discussed many of the tasks related to preparation for the evaluation design. The design can be thought of as an articulation of the plan for implementing the evaluation that typically presents the following: the model on which the evaluation is based, the program's theory of change, data collection tools and protocols to be used, analysis and interpretation strategies, reporting plan, and utilization of results. Each stage of the evaluation as presented in the evaluation design should be consciously engendered.

The following sections highlight best and promising practices for engendering evaluation at the evaluation design and implementation stage. The designated evaluation team described earlier would be intimately involved in this process.

EVALUATION MODELS

Evaluation design generally begins with determining the purpose of the evaluation and what information needs it will address, which guides the selection of an appropriate evaluation model. In the evaluation world, a model is an approach to evaluation. Each model is based on a particular vision of evaluation. Models also propose particular roles for the evaluator; for example, a practitioner using a participatory evaluation approach might act as a facilitator who focuses on engaging participation at each stage of the evaluation. Evaluators consider the program context and needs when selecting a model. Organizational learning models might be chosen in cases where there is a strong desire for significant individual or team capacity building.

It is very common for an evaluation design to be based on more than one model; often, practitioners choose more than one model to guide the development of an evaluation project. For example, a program may use a Utilization-focused Model so that the evaluation is clear about the intended use for intended users as well as a Feminist Evaluation or Transformative Evaluation model to guide its attention to gender issues. Evaluators may integrate elements of multiple designs or use one model, for example, to inform design and reporting and another to inform data collection and analysis.

A USAID Operating Unit could choose to request the use of one or more particular models when crafting the SOW based on its long-term interests and criteria or particular program needs. Evaluation specialists will also bring with them expertise and success with particular models. Whether or not a particular model is proposed by the OU, the SOW should make clear that evaluation approaches must explicitly address how a gendered perspective will be integrated throughout the evaluation framework, at all stages of the evaluation.

Some models are more suited to addressing gender equality and women's empowerment (are more gender-sensitive) than others. Utilization-focused models, for example, will focus on gender equality to the extent that such information is explicitly requested by the intended users of the evaluation. Social justice focused models are more likely to frame an evaluation in ways that put equality, empowerment, and human rights concerns at the center of the evaluation. Regardless of the model that is used, it is incumbent on the evaluator or evaluation team to highlight gender and gender equality as key dimensions of interest.

It is possible to think about three primary uses of evaluation: to judge the merit or worth of a program (including the extent to which it achieved its intended purposes), to improve a program or its management, or to generate data for specific purposes (such as policy creation) (Patton, 1997, p. 76). There are numerous evaluation models that have been examined in the literature and through practice; these can be categorized in multiple ways. The table below organizes evaluation models that are particularly suited to engendering evaluation in international development programs (see Preskill, 2004 and Patton, 2008) because each allows evaluators to assess the varying experiences of and impacts on different groups of stakeholders. The key authors linked to each model are noted below, as is the key focus of the approach.

Table I: Evaluation Models Compatible with a Gender-Sensitive Approach

Approach	Key Focus	Key Questions	Selected Associated Models
Behavioral Objectives	program objectives	Is the program, product, or process achieving its objectives?	Real World Evaluation (Bamberger, Rugh, & Mabry, 2012)
Goal-Free	actual vs. intended outcomes	What are all the effects of the program including its side effects?	Goal-free (Scriven, 1991)
Responsive	responds to the information needs of various stakeholders	What does the program look like to different people?	Responsive (Stake, 1975) Stakeholder-based (Mark & Shotland, 1983)
Utilization-focused	intended uses for intended users	What are the information needs of stakeholders and how will they use findings?	Utilization-focused (Patton, 2008) Developmental Evaluation (Patton, 2011)
Participatory or Collaborative	engaging stakeholder in the evaluation process	What are the information and strategic needs of participants?	Participatory Evaluation (Whitmore, 1999; Jackson & Kassam,1998)
Social Justice/Human Rights	empowerment, improvement of conditions, self-determination, equality	What are the information needs to foster improvement in conditions and self-determination? What is the best way to address social justice concerns?	Feminist Evaluation (Brisolara, Seigart, & Sengupta, 2014; Seigart & Brisolara, 2002; See also Hay, 2012) Transformative Evaluation (Mertens, 2008)
Organizational Learning	capacity building, evaluation as a catalyst for learning	What are the information needs of individuals, teams, and organizations?	Appreciative Inquiry (Preskill & Catsambas, 2006)

All of the above models value the involvement of stakeholders in the evaluation to some degree, particularly in the design stage. Most of the models above consider project participants to be important project stakeholders. The models differ in terms of the degree to which they recommend stakeholders being involved in the actual mechanics of the evaluation. Regardless of the model, however, there are some common recommendations with respect to stakeholders pertinent to crafting an engendered evaluation:

- Ensure that selection criteria for evaluation participation are open, transparent, and impartial (ICRW, 2009) and that the evaluation reflects the diversity of stakeholder communities (women and men, key stakeholders) (UN, 2011). In prioritizing who may be involved in the design phase, it is recommended that key decision makers, leaders, power brokers, and/or credible liaisons to communities be chosen.

- When considering how participatory the evaluation should be, consider what degree of participation (or participation in what stages of the evaluation) in the evaluation is appropriate for the evaluation and the participants. Participation results in implications for the evaluation. On the positive side, it can contribute to cultural and gender relevance, increase buy-in and commitment to the program, increase credibility of the findings, and contribute to evaluation use. Engaging participants also has implications for the evaluation timeline and budget, and requires effort. There may also be negative social consequences in cases where participation is unexpected or certain groups perceive exclusion from participatory processes. From the perspective of participants, it is important to consider the intensity or extent of involvement

desired and the implications of participation (e.g., costs incurred due to transportation, missed work, risk-taking, energy expended as well as who else might be affected by these costs).

- The evaluation team should include experts in participatory processes who have strong facilitation skills. Cultural and gender factors will help elucidate power differentials; these should be accounted for in communication strategies, the selection of decision-making processes, the location of meetings, and whether women and men or different groups should meet together or in smaller or separate formations.

EVALUATION QUESTIONS

The model(s) chosen will guide the framing of the key evaluation questions (though the questions themselves stem from the purpose of the evaluation). Evaluation questions are often informed by a focus on differential impacts for various actors as the result of the program: who gets what, who owns what, who does what, and who is affected (Ulrich, 1983). If the evaluand is a gender equality or female empowerment program, the question(s) to be answered will certainly focus on gender differences in outcomes. However, if the program has other primary goals, it is more likely that gender outcomes will be sub-questions stemming from the overarching evaluation questions.[2] As a result, one additional method for assessing the degree to which the key evaluation questions are engendered is to consider if any of the following undesirable conditions exist (Kyomuhendo, 2010):

- Does the question exclude one sex (implicitly or explicitly)?

- Does the question take the male as the norm for both sexes (implicitly or explicitly)?

- Does the question assume that men and women are homogeneous groups?

- Does the question construct men as actors and women as acted upon?

- If only one sex is being considered, are conclusions likely to be drawn in general terms with one sex being taken as the norm or standard?

Other issues to consider in reviewing evaluation questions include the following:

- Do evaluation questions point to outcomes focused on cultural and gender-specific results?

- Should evaluation questions address the need for changes in community norms or in particular attitudinal or behavioral change? If so, how?

- Are program outcomes in keeping with the Agency's goals of gender equality and female empowerment?

- Is at least one evaluation question dedicated to assessing unintended consequences?

- Does the question take the family or household as the basic analytical unit? How does this affect who is visible and invisible and what dynamics might be more difficult to perceive?

THE PROGRAM'S THEORY OF CHANGE

Articulating a program's theory of change serves many functions. A theory of change reflects a program's priorities, values, and intended outcomes; clarity about the theory of change can set agendas and frame conversations. A theory of change is an explicit statement of how an intervention is understood to contribute to the intended and observed outcome. It can also be thought of as an

---

2. Even in evaluations focused on programs other than gender equality and female empowerment, one or more of the main evaluation questions may be related to gender. Whether the gender-related questions are main questions or sub-questions depends on what information is needed and what will be done with the information produced through the evaluation.

articulation of the processes by which change is said to come about for an individual organization or a community.

Typically, individual theories of change are articulated through log frames, logic models, or results chains (Rogers & Hummelbrunner, 2012). The purpose of such "tools" is to synthesize and (often) offer a visual representation of how inputs and activities are supposed to lead to outcomes. Important recommendations offered by development practitioners in recent years include the following:

- Attempt to include the complexity and nonlinear nature of program experience.

- Reflect, as much as possible, the different perspectives of various stakeholders, acknowledging that outcomes that are positive for some may be negative for others.

- Include in the theory of change, or model, possible alternative explanations, such as what else might be contributing to the observed or desired changes (Rugh & Bamberger, 2014).

- Ground the theory of change and any visual or other interpretations/summaries in the program context. The gender analysis will elucidate key elements of the context. Other contextual factors that need to be taken into account are economic, political, legal, administrative, and environmental factors, especially those that affect the use of services by vulnerable populations (Bamberger & Segone, 2011) or that could limit gender-equality outcomes.

- Explicitly examine the extent to which the theory of change incorporates gender equality and female empowerment outcomes.

It is important to remain aware that the theory of change as well as the evaluation questions and design are a product of, and reflect the values driving, the evaluation (Mertens, 2014). Similarly, evaluation theory, evaluation models, and theories of change are products of particular contexts rooted in particular cultures (Kirkhart, 2010); understanding the values espoused explicitly and implicitly by models and theories is important in being able to identify key evaluation gender issues.

DATA COLLECTION: TOOLS, PROTOCOLS, AND METHODS

A great deal has been written on feminist (gender-focused) research theory and methods (see Brisolara, Seigart, & Sengupta, 2014; Hesse-Bieber, 2014). Emerging from these studies has been the insight that there are not necessarily exclusively gendered methods, but that practitioners should focus on how, with whom, in what situations, and why methodological choices are made. Key considerations related to these questions are listed below:

- Given the importance of participation of a diversity of participants, it is important to pretest instruments with different gender subgroups.

- Mixed-method designs, those that use quantitative and qualitative data, are strongly recommended for engendered evaluations. Reasons for including quantitative data include gathering data from a larger number of respondents, the ability to gather data from comparison or control groups, the credibility of this evidence to certain stakeholders, and the efficiency of data collection. Being able to show statistically significant differences by sex, gender identity, or other respondent traits contributes to engendering an evaluation. Reasons for gathering qualitative data include the need for understanding complex dynamics, the need for flexibility in addressing emerging issues, the nature of cultural and gender specific issues, the credibility of this evidence to certain stakeholders, and the importance of understanding unintended outcomes. Qualitative data are also frequently better for capturing the differences in how respondents experience gender differences – the language itself used by respondents can be very revealing. Collecting both quantitative and qualitative data allows for triangulation of data sources and, thus, greater confidence in evaluation findings.

- Data collection methods should be culturally appropriate. For example, a tool might use locally recognized symbols or materials, reflect or recognize cultural values (e.g., a focus group might not be appropriate in a culture where it is proper to defer to the eldest participant or to be perceived as agreeable).

- Data collection tools should be rooted in an understanding of gender roles and responsibilities within their cultural contexts. This might result in the inclusion of particular questions in instruments, the use of particular indicators, a difference in how terms are operationalized, and in selecting respondents.

- Data collection protocols need to acknowledge existing gender relations within a particular context. What interviewer-interviewee configurations will feel safest, or result in trustworthy data, will vary by context. In some cases, consultation with religious figures, male spouses, or elders might have to occur before permission to participate in data collection can be secured for women or girls (World Bank, 2005). In crafting protocols, it is important to consider accommodation issues, particularly for the most vulnerable participants. Will respondents need childcare? What childcare options are acceptable? Will respondents need interpreters (e.g., sign language, local languages)? Are there physical ability issues or safety concerns? To what extent does the way in which people are invited to participate matter? Attention to such issues must be addressed in data collection protocols.

- Bringing a gendered lens to particular data collection tools (e.g., surveys, interview guides, observation sheets) requires asking questions about who benefits and is disadvantaged within the program, in what ways does gender shape the ability to achieve program outcomes, and what gender differences exist in outcomes, experiences, and unanticipated consequences.

ANALYSIS AND INTERPRETATION

Bringing a specific gender/feminist lens into the analysis of findings as well as the development of recommendations is important in achieving gender equality (Node, 2005). Participation of people with divergent perspectives in the analysis is particularly important in avoiding interpretation that draws from stereotypes (Kirkhart, 2010). Adopting a gender or feminist lens prepares participants to examine the extent to which data is analyzed with engendered evaluation questions and goals in mind. It is not so much that there exist gendered analytical strategies, but that analysis strategies can become engendered through the application of a gendered lens or approach. Typically, this involves making the insights mentioned above explicit by questioning assumptions and asking who is benefiting and who is disadvantaged within a given situation, how (gendered and cultured) power dynamics are defining possibilities, and who is engaged in responding to social inequalities (Brisolara, Seigart, & Sengupta, 2014). The analysis process would use these questions in reflecting on the data and would apply them to the findings in order to reach deeper, gender-sensitive understandings and conclusions. The table below provides examples of questions that orient the evaluator to applying a feminist or gendered approach to data analysis (Brisolara, 2014).

Examining Gender and Social Justice with a Gendered Lens
What are the prevalent social issues and cultural values for various stakeholders and how does this help us understanding the findings? In what ways are women, men, lesbian, gay, bisexual, transgendered, and intersex people treated differently within the program and how do their experiences and outcomes differ? According to the findings, in what ways do class, race, and gender combine to expand or contract possibilities for participants? How do perspectives that account for such differences illuminate different program dynamics? What structural and gender inequalities exist within this context and how do these frame our understanding of the findings? What is gained and lost (for women, for men, for gender relationships) by acting? By not acting? What are the personal, social, and political consequences of these inequalities? What are the consequences of bringing systemic and structural inequalities to light? What are evaluation participants' most pressing needs for action, according to them?

REPORTING AND UTILIZATION

While it sounds dire to say that we forget reporting and utilization at our own peril, it is true. The primary justifications for investing time and energy in evaluations are accountability and also, importantly, contribution to the effectiveness, efficiency, and value of current and future inquiry efforts. In order to accomplish these goals, we need to concentrate, from the beginning through the end, on how the evaluation and its results will be used. The table below highlights examples of questions that may need to be considered in the presentation and dissemination of evaluation results (Brisolara, 2014).

Examining the Nature of Knowledge Dissemination and Reporting with a Gendered Lens
In what ways are the types of knowledge of interest already utilized by different groups of stakeholders? What are the consequences of sharing/not sharing what is learned for women and for men? For whom are we producing "knowledge" and for what purposes? Which forms of knowledge have the highest credibility (and does this depend on the source of information)? What personal political stances might interfere with an ability to see or represent project politics and what steps can be taken to mitigate this effect?

Evaluators are well advised to present findings with an eye to what is actionable (USAID, 2011). Reporting and utilization considerations can be thought of as questions about people and groups, forms, and utilization focus. The following considerations are drawn from multiple sources, the author's experiences, and the following sources: USAID, 2011; Bamberger, 2012; DFID, 2012; Kyomuhendo, 2010; Patton, 2011; UNEG, 2011.

PEOPLE AND GROUPS

- Considering the intended uses and users previously determined, who should receive some form of the report? Think about direct and indirect users of this information and the message that is being sent by involving or not involving particular groups of people.

- What are the information needs and interests of different audiences? Have the particular issues that are salient to different groups of women and men been addressed?

- What background information must be included in order for users to determine the credibility of the evidence provided and its significance in explaining gender-sensitive results? One example might be transparency about the sampling strategy used, who is not represented in the sample, and why.

- What information must be included in order to meet the goals of the evaluation and to adhere to the gender-sensitive evaluation criteria chosen?

- Are any measures needed to safeguard the interests of vulnerable/marginalized groups currently, after program completion, or upon dissemination of the report? The gender context must be taken into account when considering vulnerabilities and marginalization.

FORMS AND FORMATS

- Considering the different audiences and stakeholders, who will be using results and what formats are appropriate for each group? Written reports might not be the most appropriate, useful, or credible. Alternate forms should be culturally appropriate and sensitive to gender as well as other identity dimensions. Examples of alternative reporting forms include drama, dance, video, visual representation, and oral presentation.

- In what language(s) should reporting be conducted? Given credibility concerns, access issues, and utilization interests, who should present the findings?

- Are the conclusions made in different formats adequately supported by the findings?

OTHER UTILIZATION ISSUES

- Do reporting forms include conclusions that are actionable? Are actionable items articulated in a gender specific form?

- In what form will actionable items be presented: as conclusions, recommendations, or as an action plan? What effect might different forms have on the possibility that gender-focused needs will be addressed?

- Who/which organizations need to be involved in action items in order for change to occur or be sustained? Are there specific gender-focused organizations that should be consulted in order to increase attention to and work strategically on these issues?

- Do actionable items need to be organized by entry point (e.g., program, policy, or legislation) in order to be most effective? Is there a difference in language that should be considered when addressing gender-focused items depending on the entry point in this cultural context?

V. SUMMARY

We have learned a great deal from practitioners, development professionals, evaluation specialists, and reviews of evaluations about how to conduct gender-sensitive evaluations. This document highlights the key issues to keep in mind in conducting an engendered evaluation as well as strategies that have proven

successful. The literature highlights the critical roles of specificity (of sector-focused indicators and clear key terms), participation and inclusiveness (of key stakeholders and perspectives), strategies for applying a gender(ed) perspective (and asking engendered questions), cultural competence and gender expertise (rooted in local contexts), and an integrated focus on utilization or action in fostering a successful engendered evaluation. Key lessons of this review have been summarized in checklists that can be reviewed at each stage of the evaluation; these appear in Annex A of this volume to provide additional support to those in the field dedicated to this important work.

ANNEX A: GENDER RESPONSIVE EVALUATION CHECKLISTS

The following questions serve as guides in determining issues that evaluation team members should address when implementing a gender responsive evaluation. Each set of questions relates to a stage of the evaluation project and can be used to determine the extent to which the evaluation has been engendered at each stage. In using the checklists, it is helpful to describe the extent to which a particular condition has been achieved and what other factors need to be considered in further developing the evaluation. Note that the term "program" in the tables below can be substituted with project or initiative.

Questions	Yes, No, Partially	Description
Evaluation Preparation		
Do the program's goals and objectives reflect gender equity aims? Are gender equity aims well-defined? agreed upon?		

Are relevant data related to the gender equity aims of the program available?
     Are program and organization data disaggregated by sex? By race/ ethnicity, socio-economic characteristics, and/or other important demographic variables?
     Can missing/needed (baseline and outcome) data be gathered/made available for a reasonable cost?		

Does the Scope of Work (SOW) clearly state the intended gender equity uses of evaluation findings?
     Have relevant subgroups of women and men been identified? (e.g., by ethnicity, age, language, ability, region, by marital, legal, migration status)
     Have forms of inclusion of attention to these subgroups been addressed in the SOW?		

Has a gender analysis of the program's social context been completed?
     Is the analysis recent enough to serve the evaluation?
     Does the gender analysis address gender relationships? Gender roles? The sexual division of labor? Power differentials? Household decision-making? Access of public spaces? Control over one's own body? Opportunities?		

Have evaluation team members been trained in gender sensitive approaches and the importance of engendering evaluation? Does the team include gender experts? Cultural expertise? Methodological expertise? Both men and women? Representatives of key stakeholder groups?		
Are supervisors and staff aware and supportive of the mission's gender equity and empowerment aims? Have they participated in gender equity training?		
Does the current monitoring and evaluation system (and budgeting) support gender equity information needs?		

Questions	Yes, No, Partially	Description
Program Design and Outcomes		

Are the program's outcomes consistent with the gender equity aims of the program?
     Do program activities and outcomes reflect the gender needs of men and women?
     Do program activities and outcomes reflect the gender needs of important subgroups of women and men?		

Do the proposed outcomes address (or acknowledge) the (cultural, social, political) barriers to gender equity in the current context? Are there other specific barriers or challenges that the program is designed to address?		

Is the program focused on a few key outcome indicators? Do these indicators reflect gender needs and differences?
     Are the indicators culturally sensitive and appropriate?
     Are the indicators specific to the sector in which the program is situated?		

Were gender experts involved in the design of the program activities and/or outcomes?		
Do program activities address the participation needs (requirements) of both men and women? (e.g. childcare, transportation, income earning activities, etc.)		

Are outcomes. . .
     clear in general and with respect to gender difference?
     relevant to the subject at hand and gender equity aims?
     economical, available at a reasonable cost considering gender responsive evaluation needs?
     adequate for the evaluation's gender equity needs?
     measureable, both the indicator and the specific gender responsive needs?		

Has the program's theory of change been described through a logic model, log frame, narrative, or some other model?		
Have comparison groups and/or gender equity indicators (perhaps at a national or regional level) been identified and made available?		

Questions	Yes, No, Partially	Description
Evaluation Model and Participant Involvement		
Is the evaluation model(s) appropriate for unearthing the gender equity needs of the project? Is the evaluation team familiar with the particular evaluation model or approach?		
Has the role of the evaluator and evaluation team, especially with respect to gender equity and empowerment needs, been made explicit? Does the evaluation insure the active and meaningful participation of women, men, boys, girls, and people with other sexual identities?		
Does the evaluation model contribute to finding potential solutions to gender disparities?		
Does the evaluation model need to be adapted in order to effectively recognize gender equity issues, needs, and dynamics?		

Are proposed forms of stakeholder involvement consistent with the gender equity needs of the program?
    Is there adequate gender representation among involved stakeholders?
    Are methods for selecting stakeholder transparent? Impartial? Reflective of the diversity within potential participations (and representative of subgroups)?
    Are power brokers or culturally designated gatekeepers involved in participant selection?
    Has the degree of participation been negotiated with stakeholders/participants?
    Have the potential implications and/or social costs of participation (or nonparticipation) been determined? Discussed with participants?		

Do plans for stakeholder involvement reflect the relevant diversity dimensions of stakeholders? (e.g., class, race, ethnicity, age, marital status, head of household status, language, ability, sexual orientation, migration status, literacy and education level, religious or political affiliation)		
Are there male and female roles, responsibilities, or activities that could prevent participation in the program or evaluation? Facilitate participation? If so, have the program and evaluation addressed these issues in order to maximize participation?		
Will stakeholders be involved in selecting indicators? In other stages of the evaluation?		

Questions	Yes, No, Partially	Description
Evaluation Design		
Are overarching evaluation questions inclusive of both/all sexes?		
Do questions assume one sex as the norm?		
Are both women and men constructed as actors rather than as acted upon?		
Do questions address gender equity and empowerment aims? Needed changes in community norms, attitudes, behaviors, and policies?		
Do questions include a question regarding implementation of program activities? A question about unintended consequences?		
Is the scope of the design sufficient for the information needs of the program, Operating Unit, and stakeholders? (e.g., the intended use by intended users)		
Is the scope of the design realistic given the timeline and resources of the evaluation?		

Questions	Yes, No, Partially	Description
Program Theory		
Does program theory describe explicit gender equity and empowerment aims?		
Are the potential factors (economic, political, legal, environmental, cultural) that can limit gender aims taken into account?		
Does program theory address visible, hidden, and invisible power?		
Are different perspectives of diverse stakeholders reflected in the program theory?		
Does the program theory assume linear change?		
Are cultural values reflected/embedded in program theory?		
Are program (and Operating Unit) staff familiar with the program theory as described, particularly with respect to gender equity provisions of particular outcomes?		

Questions	Yes, No, Partially	Description
Data Collection		
Are both quantitative and qualitative methods included in the evaluation design?		
Are data collection methods culturally appropriate and valid?		
Do data collection tools reflect culturally specific gender roles especially within the sector of interest? Power differentials? (e.g., do tools and protocols acknowledge the different roles that men and women play in the program sector?)		
Do data collection tools reflect gender differences in cultural and sector specific knowledge held?		
Do data collection instruments contain both gender specific questions and questions common to all?		
Do protocols acknowledge gender realities? (e.g., culturally appropriate forms of gaining access to respondents, who should interview whom)		
What other needs must be addressed to allow both women and men to participate in data collection efforts? (e.g., transportation needs, stipends, child or elder care duties, interpreters, safety needs)		
Are there laws, policies, and institutional practices that may contain implicit or explicit gender biases and that may need to be addressed by the project? Which are they?		

Questions	Yes, No, Partially	Description
Analysis and Interpretation		
Does the analysis disaggregate data by sex/gender?		
Have relevant experts been included in analysis and interpretation? (e.g., gender, cultural, methodological, various stakeholder perspectives)		
Is it clear whose voices and perspectives are included, excluded, diminished in the analysis? To what extent have gender relations or gendered power dynamics shaped this situation?		
Have participants been treated differently as the result of their sex, gender identity, or sexual orientation?		
Have participants experienced different outcomes as the result of their sex, gender identity, or sexual orientation? Does the analysis address how sex and gender have interacted with class, race, ethnicity, ability, and other important dimensions to affect results and/or program experiences?		
Does the analysis present a description of the gendered context in which the program is situated? (e.g., the gendered division of labor, access to, control of, and use of resources, household decision making)		
Has the analysis addressed the structural gender inequities in this context? The consequences of bringing (and not bringing) such inequities to light?		
Does the interpretation include a discussion of the personal, social, and political consequences of existing inequities? Have local gender experts and organizations been able to review or comment on initial findings or interpretations?		
Have unintended consequences been analyzed?		
Have the sampling strategies, data collection protocols, and other important evaluation decisions been described (made transparent)?		
Does the analysis triangulate data from different sources?		

Questions	Yes, No, Partially	Description
Dissemination and Reporting		

Has the evaluation team, in conjunction with others, created an action or use plan based on findings and further information needs?
     Does this action plan directly address gender equity needs?
     Does the report and action plan describe participants' needs, including gender needs, according to the stakeholders' themselves?
     Does the action plan address what is gained and lost by acting on findings? By not acting on findings?
     Do action plans and/or recommendations address program, organizational, community, policy, and legal gender equity issues?		

Is (are) the reporting or dissemination form(s) appropriate for the intended audiences? Are they credible? Are additional formats needed? Are reports equally accessible to women and men? By vulnerable or marginalized populations?		
Do the report and dissemination strategies reflect an awareness of the safety needs of participants and stakeholders, particularly the most vulnerable or marginalized groups?		
Do the reporting formats report on data in disaggregated form?		
Does the scope of reporting and dissemination meet the information needs of intended users?		
Is additional information needed in order to assist users in meeting the gender equity aims of the program?		
Are local or regional institutions, departments and journals that focus on gender included among the target groups for dissemination, along with mainstream research magazines? Are there specific publications or events for gender-related findings that should be considered?		
Are the contributions of various stakeholders to the program's success acknowledged?		

REFERENCES

African Development Bank. (2011 May). *Mainstreaming gender equality: a road to results or a road to nowhere? An Evaluation Synthesis*. ADB Operations Evaluation Department.

American Evaluation Association's Guiding Principles for Evaluators, http://www.eval.org/p/cm/ld/fid=51

Bamberger, M. & Segone, M. (2011). *How to design and manage equity focused evaluations*. New York, NY: UNICEF.

Bamberger, M. (2013). *Engendering M&E*. [PREM Notes number 17]. Washington, DC: World Bank.

Bamberger, M., Rugh, J. & Mabry, L. (2012). *Real world evaluation: working under budget, time, data, and political constraints*. Los Angeles, CA: Sage.

Bradshaw, S. and Linneker, B. (2003). *Challenging women's poverty: perspectives on gender and poverty reduction strategies from Nicaragua and Honduras*. London: Catholic Institute for International Relations.

Brisolara, S. (2014). Feminist theory: its domains and applications. In S. Brisolara, D. Seigart, & S. Sengupta (Eds.), *Feminist evaluation and research: theory and practice*. New York, NY: Guilford Press.

Brisolara, S., Seigart, D., & Sengupta, S. (2014). *Feminist evaluation and research: theory and practice*. New York, NY: Guilford Press.

Bustelo, M. (2011). *Last but not least: gender-sensitive evaluation as a forgotten piece of the policymaking process*. Unpublished manuscript.

CARE (2012). "Good practices framework: gender analysis." CARE International Gender Framework.

CIDA (1997). *Guide to gender sensitive indicators*. Government of Canada.

CIDA (2006). *Gender based analysis policy*, Indian and Northern Affairs.

CIDA (2010). *CIDA's policy on gender equality, policy and tools*. Quebec: CIDA.

Corner, L. (2003). *From Margins to Mainstream: From Gender Statistics to Engendering Statistical System.*, New York, NY: UNIFEM.

De Waal, M. (2006). Evaluating gender mainstreaming in development practices. *Development in Practice, 16*(2), 209-214.

Dunaway, K.E., Morrow, J.A., and Porter, B.E. (2012). "Development and validation of the cultural competence of program evaluators (CCPE) self-report scale." *American Journal of Evaluation, 33*(4), 496-514.

Espeln, A. and Brody, A. (2006). *Putting gender back in the picture: rethinking women's economic empowerment.* Report prepared at the request of the Swedish International Development Cooperation Agency (Sida)**.**

Espinosa, J. (2013). "Moving towards gender-sensitive evaluation? Practices and challenges in international development evaluation." *Evaluation, 19*(2), 171-182

Espinosa, J. (2011). *La igualdad de género en la evaluación de la ayuda de desarrollo: los casos de la cooperación official británica, sueca y española.* Unpublished doctoral thesis, Universidad Complutense de Madrid.

FAO (2011). "Evaluation of FAO's role and work related to gender and development." Office of Evaluation (OED) June.

Gender and Development Network (2012). *A theory of change for tackling violence against women.* DFID. How to note series, guidance note 1.

Gender, Agriculture, & Assets Project, GAAP (2012). *GAAP: A toolkit on collecting gender and assets data in qualitative and quantitative program evaluation.*

Golla, Malhotra, Nanda, and Mehra (2011). Understanding and measuring women's economic empowerment: definition, framework and indicators. ICRW.

Hay, K. (2012). In *"Evaluation for equitable development results"*, 2012, UNICEF in partnership with Coneval, IDEAS, IDRC, ILO, IOCE, UNDP and UN Women.

Hesse-Beiber, S (Ed.) (2012). *Handbook of feminist research theory and praxis*, 2nd ed. Los Angeles, CA: Sage.

Hesse-Bieber, S. (Ed.) (2014). *Feminist research practice: a primer*, 2nd ed. Los Angeles, CA: Sage.

ICRW (2009). "Innovation for women's empowerment and gender equality", with Anju Malhotra, Jennifer Schulte, Payal Patel, and Patti Petesch.

ICRW (2013) Cultivating women's participation: Strategies for gender-responsive agriculture programming: Tanzania gender and agricultural forum.

IFPRI & ILRI (2012) *A toolkit on collecting gender & assets data in qualitative & quantitative program evaluations.*

ILO (2007) *A manual for gender audit facilitators: The ILO participatory gender audit methodology.* Geneva: International Labour Organization.

Kabeer, N. (1994). *Reversed realities: gender hierarchies in development thought.* London, UK: Verso.

Kanesathasan, A, Jacobs, K., Young, M., and Shetty, A. (2013). *Capturing the gender effect: guidance for gender measurement in agriculture programs.* Technical Brief. Washington, D.C.: International Center for Research on Women.

Kanesathasan, A. (2012). *Cultivating women's participation: strategies for gender-responsive agriculture programming.* Technical Brief. Washington, D.C.: International Center for Research on Women.

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

Kosheleva, N. & Segone, M. (2013). EvalPartners: Facilitating the development of a new model of voluntary organization for professional evaluation to support the development of national evaluation capacities. *American Journal of Evaluation*, *34* (4), 568-572.

Kusek, J.Z. & Rist, R.C. (2004). "Ten Steps to a Results-based monitoring and evaluation system."

Kyomuhendo, G.B. (2010). "Gender focused methods of research." Naivasha, Kenya: Makerere University, IDRC.

Lexow, A.S.J. & Hansen, S. (2005 July). *Gender and development: a review of evaluation reports 1997-2004*. Oslo: Nordic Consulting Group.

Mark, M.M. & Shotland, R.L. (1985). "Stakeholder-based evaluation and value judgments." *Evaluation Review, 9*(5), 605-626.

Mertens, D. (2009). *Transformative research and evaluation*. New York, NY: Guilford.

Mertens, D. (2014). Enabling gender responsive and equity focused evaluation. Webinar through Eval partners, January 30th, 2014.

Monbart, C. & Rathner, M. (2010 October). *Evaluation of strategic programme objective 7: Enhancing research-policy linkages on social transformations*. New York, NY: UNESCO.

Morra-Imas, L.G. & Rist, R. (2009). "The road to result: designing and conducting effective developmental evaluations." Washington, DC: World Bank.

Nordic Development Fund, NDF (2004). *Gender equality study: a study on addressing gender aspects in projects co-financed by NDF in developing countries*. March. Helsinki: NDF.

NIBR (2006). *Lessons from evaluations of women and gender equality in development cooperation.* Oslo: Norwegian Institute for Urban and Regional Research (NIBR), October.

OECD (2010). The OECD Social Institutions and Gender Index. http://www.oecd.org/dev/poverty/theoecdsocialinstitutionsandgenderindex.htm

Oxfam (2004). *See both sides: a practical guide to gender analysis for quality service delivery.*

OSCE (2011). The Secretary general's annual evaluation report on the implementation of the 2004 OSCE *Action Plan for the Promotion of Gender Equality Organization for Security and Co-operation in Europe,* September**.**

Paluck, E. & Ball, L. (2010). "Social norms marketing aimed at gender based violence: a literature review and critical assessment*."* International Rescue Committee**.**

Patton, M. Q. (1997). *Utilization-focused evaluation*. St. Paul, MN: Sage.

Patton, M. Q. (2008). *Utilization-focused evaluation, 4th ed*. St. Paul, MN: Sage.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.

Patton, M.Q. (2011). Developmental evaluation for equity-focused evaluations. In M. Segone (Ed.) *Evaluation for equitable development results*. New York, NY: UNICEF.

Preskill, H. & Catsambas, T. (2006). *Reframing evaluation through appreciative inquiry*. Thousand Oaks, CA: Sage Publications.

Reinharz, S. (1992). *Feminist methods in social research*. New York, NY: Oxford University Press.

Rogers, P. & Hummelbrunner, R. (2011). Methodological Challenges in using programme theory to evaluate pro-poor and equity-focused programmes. In M. Segone (Ed.) *Evaluation for equitable development results*. New York, NY: UNICEF.

Rossi, P.H., Lipsey, M. & Freeman. H.E. (2004). *Evaluation: A systematic approach*, 7th ed. Thousand Oaks: Sage Publications.

Rugh, J. & Bamberger, M. (2014) Real World Evaluations. Conducting equity-focused and gender-responsive quality evaluation under constraints. EvalPartners e-learning series on Equity-focused and gender-responsive evaluations.

Ruxton, S. (ed.) (2004). Gender equality and men: learning from practice. Banbury Road: Oxfam GB.

Scriven, M. (1991). Pros and cons about goal-free evaluation. *American Journal of Evaluation*, 12, 55-62.

Seigart, D. and Brisolara, S. (Eds.) (2002). Feminist evaluation: Explorations and experiences. *New Directions for Evaluation*, 96.

Shadish, W. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19 (1), 1-19.

Sielbeck-Bowen, KA, Brisolara, S., Seigart, D. Tischler, C., and Whitmore, E. (2002). "Exploring feminist evaluation: the ground from which we rise." *New Directions for Evaluation*, 96, 3-8.

Stake, R.E. (1975). "Program evaluation particularly responsive evaluation." Center for Instructional Research and Curriculum Evaluation. Urbana-Champlain: University of Illinois.

Ulrich, W. (1983). Critical heuristics of social planning: a new approach to practical philosophy. Stuttgart (Chichester): Haupt.

UN AIDS (2012). Agenda for accelerate country action for women, girls, gender equality and HIV: Operational plan for the UNAIDS action framework: addressing women, girls, gender equality and HIV UNAIDS. Joint United Nations programme on HIV/AIDS.

UN HABITAT (2012). *Gender issue guide: Housing and slum upgrading.* Nairobi, Kenya: UN-Habitat.

UNDP (2009). Gender Equality Institutional Assessment (GEIA). Updated June.

UNEG (2011). "Integrating Human Rights and Gender Equality in Evaluation: Towards UNEG Guidance. United Nations Evaluation Group Human Rights and Gender Equality Task Force.

USAID (2011). *Evaluation: learning from experience.* [USAID Evaluation Policy]. Washington, DC: USAID.

USAID (2012a). *Gender Equality and Female Empowerment Policy.* Washington, DC: USAID.

USAID (2012b). *Toward Gender Equality in Europe and Eurasia: A Toolkit for Analysis.* Washington, DC: USAID.

Vainio-Mattila, A. (2010). Navigating gender: a framework and a tool for participatory development. Helsinki, Finland: Ministry for Foreign Affairs, Department for International Development Cooperation.

Warren, H. (2007 July). Using gender analysis frameworks: theoretical and practical reflections. *Gender and Development,* 15 (2).

Whitmore, E. (Ed) (1999). Understanding and practicing participatory evaluation. *New directions for evaluation*, 80, 5-24.

World Bank (2005). *Gender issues in the use of measurement strategies, indicators, and sampling*. Presentation by Lucia Fort and Padma Karunaratne, MENA Regional Gender Workshop Amman, May 15-19.

World Bank (2007). *Global monitoring report: confronting the challenges of gender equality and fragile states*. Washington, DC: World Bank.

World Bank (with FAO and IFAD) (2008). Gender issues in monitoring and evaluation [Module 16]. *Gender in Agriculture*. Washington, DC: World Bank.

Yellow Window Management Consultants (2009). Engender: Gender at work. *European Commission toolkit: Gender in EU-funded research*.