This article tackles this question by reflecting on the experience of the Independent Evaluation Group at the World Bank in implementing its “gender flag”—a tool developed to systematically capture gender dimensions in projects by extracting information from self-evaluation completion reports. It concludes that, despite their limitations, checklists and gender flags can be useful tools to track gender results in projects and increase evaluators’ awareness of gender and distributional impacts more broadly.
Over the past 17 years, the World Bank Group has made much progress in advancing the gender agenda. Gender is currently one of the most prominent corporate objectives. The 2015 World Bank Group gender strategy “sets more ambitious targets, establishes a new methodology for measuring progress, and outlines an agenda for new frontier areas in which our impact might be transformational. It moves away from the concept of ‘mainstreaming’ toward focusing on proven interventions that achieve tangible results” (World Bank Group 2015, 1). Achieving gender equality is recognized as a key component of eradicating poverty and attaining shared prosperity and inclusion. Its importance is reiterated under the IDA18 replenishment and the recent World Bank Group capital increase. 

How do we know if the institution is making headway in advancing gender equality in its work? In response to the renewed focus on gender, the Independent Evaluation Group (IEG) of the World Bank Group adopted a strategic plan to improve the integration of gender across its evaluation work (Box 1). The main objective of this plan is to identify viable approaches to integrate gender into the evaluation of strategies and operations systematically, so that specific gender results can be assessed and documented.

“Checklists like the gender flag adopted by IEG are useful tools to assess more systematically whether, and how well, gender issues have been addressed in country program frameworks, sector programs, or projects”.

IEG produces Implementation Completion and Results Report Review (ICCRs). These are independent, desk-based validation reports that mostly rely on the information included in Implementation Completion and Results Report (ICRs). Although gender elements are often relevant dimensions of a project, they may go

---

**BOX 1. GENDER AND THE INDEPENDENT EVALUATION GROUP’S MANDATE**

In its role of independent evaluator of World Bank Group activities, the Independent Evaluation Group (IEG) is tasked with assessing results of projects and programs, including results on closing gender gaps and assessing the effectiveness of Bank Group strategies, including its strategic approach to gender. IEG evaluates the activities of the World Bank (the International Bank for Reconstruction and Development and International Development Association), the International Finance Corporation’s support in private sector development, and Multilateral Investment Guarantee Agency guarantee projects and services “to provide an objective assessment of the results of the Bank Group’s work and to identify and disseminate lessons learned from experience.”

In 2015, IEG committed to documenting gender features in projects at the time it produces the Implementation Completion and Results Report Review—an independent, desk-based, critical validation of the evidence, content, narrative, and ratings included in the World Bank’s Implementation Completion and Results Report.

Learn more at [http://ieg.worldbankgroup.org](http://ieg.worldbankgroup.org).
How Useful are Checklists to Assess Gender in Development Projects?
unreported in the ICRR even when the ICR provides pertinent information.

This is because the current ICRR guidelines do not explicitly require evaluators to report on gender outcomes or discuss gender-relevant project’s features unless these elements are central to the operation. When the project development objectives (PDOs) have a clear gender goal—for example, increasing gender equality, closing gender gaps, or achieving specific results for women or girls—project ratings will (also) depend on gender performance. In this case, the ICR needs to provide evidence of results and the IEG evaluator can assess explicitly. But these cases are rare—less than 5 percent, according to an IEG review conducted for the Results and Performance of the World Bank Group 2015 (World Bank 2016a). Most of the time, gender aspects are “secondary” elements that may not have been fully developed in projects or may have produced unintended effects. For this reason, even when the ICR includes some discussion on gender dimensions, these results may not be recognized and reported by IEG evaluators.

IEG introduced the gender flag to systematically document the gender features of individual projects, even those that may be expected, given the nature of the project, but are not fully acknowledged in the ICR. A gender flag is essentially a checklist produced for each project-level validation report. Generated at the same time of the ICRR by the same evaluator, the gender flag is not formally part of the ICRR but is an internal supplement that aims to capture information about the project’s activities and results related to gender and identify missed opportunities, given the nature of the project.

In addition to capturing gender results, the gender flag can facilitate better reporting of distributional impacts in the ICRR. Generally, all projects have distributional impacts—whether they acknowledge it or not. Impacts may be expected to differ by gender but also by poverty level, ethnicity, geographical location, migration status, and so on. The ICR often does not report outcomes by categories of beneficiaries, but as general averages that are completely uninformative of distributional impacts. The gender flag may facilitate a more comprehensive and critical reporting of distributional impacts in the ICRR, whereby the evaluator is able to garner evidence of differential impacts for different groups when it exists or to comment on its absence.

Improving gender information in ICRRs is fundamental as ICRRs are the building blocks of many other IEG products. By ensuring more systematic reporting on gender results in validation reports, other evaluation products, such as thematic evaluations, country evaluations, synthesis papers, and the Annual Results and Performance Report, benefit as well.

What Useful Information Can a Gender Flag Provide?

The gender flag organizes any information on gender provided by the self-evaluation report (ICR) under four main headings. It captures (i) whether gender is a relevant aspect of the PDO or of one of the project components, (ii) whether the ICR reports sex-disaggregated or female- or male-specific indicators, (iii) whether there are indicators that could have been sex-disaggregated and were not, and (iv) whether the ICR discusses other specific gender issues. Evaluators need to use their experience and exercise their individual judgment to spot omissions. They are called on to determine whether the PDO or any of the project components are missing relevant gender dimensions (either at design or restructuring). For example, a project facilitating access to finance is expected to have a
specific strategy to reach out to excluded groups—women being one of those. Also, evaluators are asked to determine whether there are indicators that could have been sex-disaggregated but were not or relevant indicators that were missing.

The IEG gender flag can help to take stock of the type of gender indicators included in the projects’ results frameworks and assess their usefulness and relevance. A review of the IEG gender flag found that (i) outcome gender indicators are much less frequent than output gender indicators; (ii) person-level indicators are not systematically sex-disaggregated, even when this could be technically possible; (iii) most gender indicators relate to women or girls only, much more rarely to men or boys and very infrequently capture gender gaps; and, (iv) the use of the female beneficiaries indicator is increasing, but this indicator is often not meaningful (for example, in cases where communities or geographical areas are targeted—such as in infrastructure projects—the percentage of female beneficiaries is often reported as 50 percent). On the positive side, the IEG review found that when gender-relevant outcomes indicators are included in results frameworks at design or restructuring, they are often reported at completion (World Bank 2016a).

The components of the flag that generate quantitative information (for example, presence or absence of gender in the PDO in project activities and in results frameworks) are useful for regular aggregated reporting. Statistics from the gender flag have been included in the past three issues of the Results and Performance of the World Bank Group, the IEG annual flagship report. Monitoring the level of gender inclusion in projects keeps the Bank Group focused on the implementation of the gender agenda.

The gender flag has also been used for major thematic evaluations and learning exercises, typically in combination with additional information. For example, the evaluations of the Bank Group’s support for water supply and sanitation services (World Bank 2016b), urban transport (World Bank 2017a), and higher education (World Bank 2017b) used the gender flag and its template to extract information on gender from the respective portfolios. The gender flag has also been presented and discussed in evaluators’ forums as a tool to improve the assessment of gender issues by independent evaluation offices.

Limitations of the Gender Flag

The gender flag is a standardized and limited tool to collect information and cannot respond to specific evaluation needs. Evaluation teams of major thematic evaluations will need to adapt this tool to their case and use...
more refined approaches to gather additional information. For example, an IEG evaluation of World Bank support to rural nonfarm economy used a more sophisticated screening tool to review project appraisal documents and determine (i) whether both men and women were consulted to inform project design; and (ii) whether a gender diagnosis was conducted to understand the relationships between men and women, their roles, their access to services, markets or assets, and limitations and opportunities to address these aspects (World Bank 2016c).

For closed projects that incorporated gender indicators at design, the above-mentioned evaluation also assessed the benefits accrued by women and men, by analyzing the value chain and reported social and economic outcomes. The citizen engagement evaluation analyzed how inclusive mechanisms for the engagement of women and other vulnerable groups are, by adding a screening field like the gender flag in the portfolio review template (World Bank 2018).

The fact that both the ICCRs and the gender flag rely on information from the ICR is a critical limitation - there is no option for collecting new evaluative material. Therefore, the gender flag cannot capture gender results unless these are reported by project teams. Often, however, reporting of gender results has multiple limitations. Project teams favor narrative reporting over quantitative statistics; and, there is a systematic bias in favor of success stories. Generic statements, anecdotes, and plausible (but not quantified) impacts are often included in ICRs, but these are not particularly informative.

It is important to recognize upfront that the information provided by the flag may suffer from several biases. If the flag does not identify any gender impact, it is difficult to distinguish whether this is because

- the project did not generate any (either because it was not aiming to or because it simply did not), or
- no impact was reported in the ICR, or
- the evaluator did not discern any impact (either because he or she did not pick up information present in the ICR or refrained from commenting on missed opportunities).

Spot checks are required to detect some of these biases and other validation exercises are needed to obtain more detailed data.7

Although some elements of the flag are easy to check and report on, other aspects may be challenging for evaluators, especially those who are not familiar with gender and distributional issues. For example, detecting omissions may
require knowledge of the theoretical and empirical literature on a specific topic, especially with reference to gender issues, and an understanding of contextual elements, including gender norms and roles in the specific country. Evaluators may quickly define gender aspects as “not relevant” in a given project because they are not able to fully appreciate its distributional implications. Evaluators may find the questions of the flag that require his or her subjective assessment too daunting and skip them altogether. Evaluators sometimes misinterpret the essence of “gender information” and equate, for instance, “gender” with “women,” not recognizing that gender has to do with gender relations and gender gaps. Missing information in the gender flag, however, may be due to the poor quality of the ICR, and not necessarily due to the poor attention or gender skills of the evaluator.

Implementation Challenges

Because 400 ICRRs are produced annually, ensuring good implementation of the gender flag is no easy task. Communication, training, monitoring, and evaluation of the gender flag are activities that require time and resources but are essential to produce reliable information.

The gender flag relies on the ability (and willingness) of the evaluator to assess and report the gender information included in the ICR, comment on the appropriateness of the chosen indicators, and signal omissions. Our reviews of the gender flag suggest that evaluators have difficulties, particularly in identifying how relevant gender is to a project (for the PDO, the project activities, the indicators) and are reluctant to make a judgment call; therefore, missed opportunities are frequently not reported in the gender flag. Training on the gender flag (currently delivered at every ICRR training session) is essential to provide examples and increase the confidence of evaluators.

Monitoring of the gender flag is challenging. The gender flag is not part of the formal validation process and is not checked during the standard ICRR quality assurance process; the IEG gender team is not able to ensure regular monitoring. As a result, evaluators may not get the immediate feedback they need to make sure they are accurately capturing gender information.

Conclusion: More Attention to Gender Indicators During Design Is Essential to Capture Gender Results

Evaluation offices are increasingly keen on mainstreaming gender in evaluation rather than producing occasional assessments of the implementation of gender strategies. Checklists like the gender flag adopted by IEG are useful tools to assess more systematically whether, and how well, gender issues have been addressed in country program frameworks, sector programs, or projects (ECG 2017). This is a first step to allow for regular harvesting of gender information, which can be fed across evaluation products.

A gender flag that captures sex-disaggregated and gender-relevant indicators in self-evaluation reports can be very helpful in advancing the discussion of what good indicators for measuring gender results are. Operational teams and monitoring and evaluation specialists are constantly struggling for adequate measures of gender gaps and gender equality. Reflecting on what we currently measure is an important first step to understand how to improve. This is particularly critical outside of human development sectors, such as education or health, where traditionally there has
been more attention to measure gender dimensions. Infrastructure projects, for example, are often perceived to be benefiting the whole community, with little asked about who may be benefiting more, or differently, than others. Some level of engagement with operational teams could help improve the currently used indicators at the stage when results frameworks are designed—a very concrete and useful contribution evaluation should aspire to make.

Endnotes


2. Gender is a special theme of the International Development Association (IDA) 18 replenishment. Under IDA18, the World Bank Group commits to target remaining first-generation gaps in human endowments, in particular maternal mortality and gaps in secondary educational attainment. Policy actions will also help remove constraints for more and better jobs for women; support women’s increased access to financial services and control over assets; build the evidence base and increase operations that address issues of voice and agency, especially in fragile countries; and enable country-level action for more and better data and evidence (IDA 2016).


4. For the preparation of the Implementation Completion and Results Report (ICR) Review, evaluators also refer to the project appraisal document, the legal agreement, the documentation related to restructuring (if applicable), and the relevant country assistance strategies. The ICR is, however, the chief source of evidence.

5. The gender analysis conducted for the Results and Performance of the World Bank Group 2015 distinguished between projects that may have the potential to positively influence gender inequalities and biases, those that may have the potential to damage gender relationships or worsen biases, and those that may take advantage of behavioral differences between men and women to amplify the project’s impact. At the design stage, projects may take advantage of these channels for gender impacts and integrate some activities or indicators. Or projects may ignore these potential effects, but gender impacts may be produced anyway.

6. Person-level indicators are indicators that can be meaningfully collected for a single individual (such as education level) as opposed to the household (such as dwelling characteristics) or the community (such as the size of the village).

7. The Independent Evaluation Group has conducted ad-hoc reviews of gender flags by checking them against the ICR. Efforts are ongoing to detect evaluator effects and find appropriate solutions, including improving the training.
References


How Useful are Checklists to Assess Gender in Development Projects?
Elena Bardasi is a Senior Economist with a background in labor, poverty, and gender. She has worked at the World Bank since 2003 and in IEG since 2012, where she recently completed the Citizen Engagement Evaluation. She was also the task-team leader of two major learning reports on the Impacts of Community Driven Development Interventions on Women’s Empowerment and Social Safety Nets and Gender, and of the thematic chapter on gender in the 2015 Results and Performance Report of the World Bank Group. As IEG thematic coordinator for gender, she has been instrumental in facilitating the integration of gender in project and thematic evaluations. She contributed to developing evaluation quality standards by coauthoring the IEG guidelines Integrating Gender into IEG Evaluation Work. Her evaluation work also includes a number of in-depth project evaluations. Elena has also written and published about female entrepreneurship, female employment and gender issues in formal and informal labor markets, wage differentials, occupational segregation, and time use.

Gisela Garcia is an Evaluation Officer in IEG’s Human Development and Economic Management Unit. Her current work focuses on assessing the gender impacts of poverty reduction policies and programs. An Evaluation Officer with 14 years of experience working on international development, Gisela has been a core team member of several major evaluations including the 2015 Results and Performance Report of the World Bank Group, the Systematic Country Diagnostic and Country Partnership Framework scd/cpf evaluation and the Citizen Engagement Evaluation. She co-authored two major learning reports on the Impacts of Community Driven Development Interventions on Women’s Empowerment and the Social Safety Nets and Gender Report. Through her gender work in IEG, Gisela has contributed to developing evaluation quality standards coauthoring Integrating Gender into IEG Evaluation Work Guidelines. Prior to joining IEG, Gisela was an Operations Officer at the Gender Group of the World Bank Poverty Reduction and Economic Management network. Before joining the World Bank Group in 2006, Gisela worked at the Evaluation Office of the Inter-American Development Bank in a variety of thematic and country evaluations, validation and evaluation reports. An Argentinean economist, Gisela holds a Master in Public Policy from The George Washington University.