

HOW IS THE UN SYSTEM REPORTING ON GENDER RESULTS: A META-ANALYSIS OF UN EVALUATIONS

UN Reporting on Gender Results

The UN has been emphasizing strengthened results reporting for at least the last decade, however reporting, including on Gender Equality and the Empowerment of Women (GEEW), often remains at an inadequate level for determining actual results achieved individually and systemwide. This is problematic on a number of levels, in particular the UN is not able to fully establish the results it has achieved and demonstrate that it is using resources as efficiently as possible, including in areas of comparative advantage.

Member States, including through the Funding Compact, and in General Assembly resolution 72/279 OP28(a) therefore requested the United Nations development system, to provide “annual reporting on system-wide support to the Sustainable Development Goals and present aggregated information on system-wide results, by 2021”. Starting this year, the UNSDG will report on the system-wide contribution/results towards the SDGs annually, as per the request of resolution 72/279 and the commitments made by the UN development system in the Funding Compact.

To promote strengthened reporting at the individual entity and system-wide levels GEEW Results UN Women has constituted a time-bound Working Group on UN GEEW Results Tracking and Reporting in collaboration with the CEB Strategic Planning Network (SPN).

UN-SWAP Results Indicators

The UN-SWAP Framework endorsed by the CEB in 2012 set out a plan for three aligned foci of accountability for gender equality and the empowerment of women, including corporate processes and institutional arrangements at the individual entity level; joint processes and institutional arrangements within the UNCT; and development results at country and normative levels.

Work that was begun under UN-SWAP 2.0 to strengthen results reporting on GEEW was put on hold during the initial phases of the COVID-19 pandemic and has been picked up in 2022. Among the tasks undertaken by the Working Group on UN GEEW Results Tracking and Reporting, a meta-analysis of UN evaluations aimed to find out how results are assessed in gender-related evaluations as well as the levels at which gender-related results are being reported along the results chain, as well as the actual results achieved.

The meta-analysis consisted of a sample of 24 evaluations purposively selected from a population of 330 evaluations on the UNEG website found after a search for the keyword “gender”. The sampling frame was a mix of the main types of gender-related evaluations on gender strategies, policies and programmes implemented in the period from 2017 to 2022 with a main focus on results. The sample is representative of the different types of entities reporting to the UN-SWAP, including evaluations at both the country and HQ levels and Joint Programme evaluations. The analysis is divided into two parts: the focus and the quality of the evaluation; the evaluation results.

Main findings

1. Evaluation focus and quality

Figure I. Evaluation focus

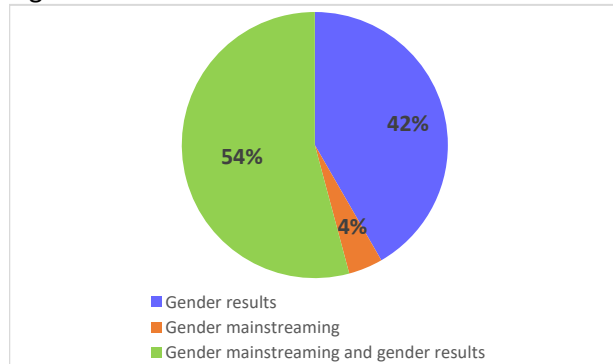
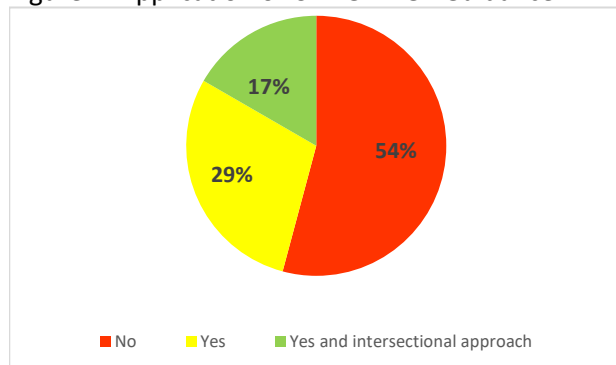


Figure I shows that just over half of the analyzed evaluations (54%) focus on gender mainstreaming and gender results, whereas the 42% focus solely on gender results.

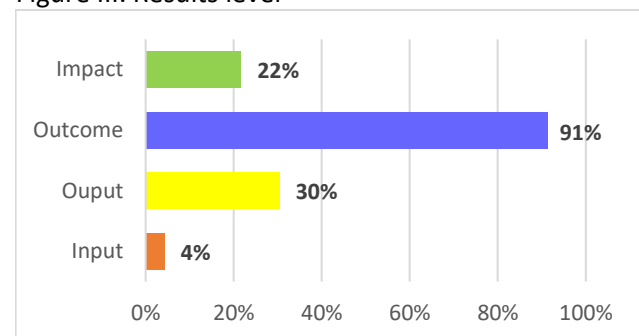
Figure II. Application of UNEG HRGE Guidance



Looking at the application of the UNEG Guidance on Integrating Human Rights and Gender Equality (HRGE) in Evaluations (Figure II), 54% of the evaluation reports did not specifically mention this Guidance. However, it must be noted that a few of them mentioned other UNEG guidelines such as the Guidance on Evaluating Gender Mainstreaming or UNEG guidelines in general, without explicitly referring to the UNEG HRGE Guidance.

While 46% of the evaluations apply the HRGE Guidance, only 17% apply an intersectional approach with a focus on Leaving No One Behind (LNOB). The intersectional approach resulted in higher attention paid to interviewing a wide range of stakeholders, to engaging with affected population; to using the right-based approach to identify and analyze inequalities and discriminatory practices; and to reaching the most vulnerable children at risk of multiple discriminations.

Figure III. Results level



The result level (activity, input, output, outcome and impact) on which the select evaluations focused was analyzed in evaluations with gender results (twenty-three out of the twenty-four reviewed evaluations). Most evaluations focus on more than one result level, which means that percentages shown in Figure III add up to more than 100%. The outcome level ranks first (present in 91% of the evaluations), followed by the output level (in 30% of the evaluations) while the impact level and the input level in the fourth place. As regards outcomes, they constitute the only level analyzed in some evaluations, while they are examined together with outputs or with impacts in others.

Limitations to reporting on results were mentioned in all analyzed evaluations. Following is a summary of the limitations most frequently reported or that require attention.

Limitations to reporting on results

Inconsistent and inadequate monitoring and results reporting systems. Lack of systematic compilation of gender-related interventions; inadequate capacity for monitoring progress with regard to gender mainstreaming; fragmented monitoring systems during the period of implementation of the strategic plan; no reporting against the entity's gender policy objectives; lack of comprehensive, department-wide tools to systematically tag content by thematic area. Evaluation reports warn that this limitation reduces the accountability of managers and the effectiveness of oversight systems, making it impossible to evaluate certain elements of effectiveness.

Lack of data. Limited sex-disaggregated and intersectional data available; lack of systematic collection of specific indicators for gender equality results or inconsistent data collected in different manner over time; data insufficiently disaggregated on the basis of geography, ethnicity and other programmatically relevant specificities.

Outcome level results. No collection of data on outcome level results (as opposed to data on activities); lack of organized data pertaining to outcome indicators at local levels; absence of a context-specific strategy that made it difficult to establish linkages between activities and outcomes achieved at country level; not possible to report on contributions at outcome level, focus on the achievement of outputs during the programme cycle.

Measurement of behavioral changes. Project and programme monitoring systems focusing mostly on activities and outputs, without information on changes in behaviors as a result of the programme inputs, make difficult to find results at the level of capacity change or behavior change. **Gender analysis** across programmes has a major focus on quantitative reporting on beneficiaries rather than also including a qualitative gender analysis.

Time constraints. Short timeline for data collection and for analyzing inconsistencies; limited time spent in country; survey instruments not pre-tested.

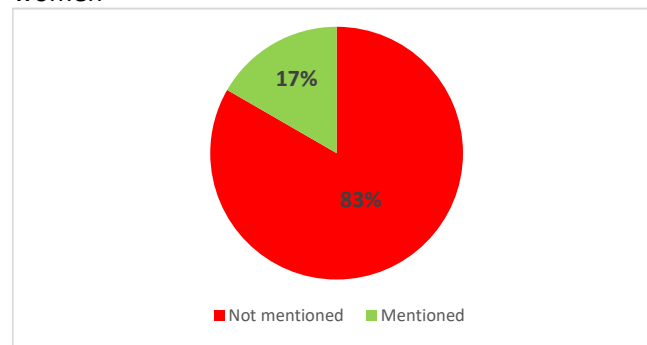
Stakeholders. No field visits and no direct consult with right holders (in some cases due to the COVID-19 pandemic), focus groups only in HQs.

Different understandings and interpretation of the term 'gender equality' across the entity, difficulties for survey respondents to understand technical gender terminology.

Unclear Terms of Reference (TOR) and Theories of Change (TOC). Unclear TOR leading to confusion on data collection and coordination; absence of TOC defining clear outcomes that linked entity's implementation mechanisms to the gender policy objectives; change from a standalone objective to a cross-cutting strategic framework theme, making it more difficult to collect data, especially in terms of gender policy results and impact. In some case there is an overlap between policy objectives, indicators in the results framework and the Gender Action Plan, which created confusion for country offices about what data should be collected and for what purpose.

Time lag. The time lag between policy development implementation and results is acknowledged as a limitation, particularly when looking at results. This has been addressed by examining trends rather than single-year data.

Figure IV. Limitations to engagement with marginalized women



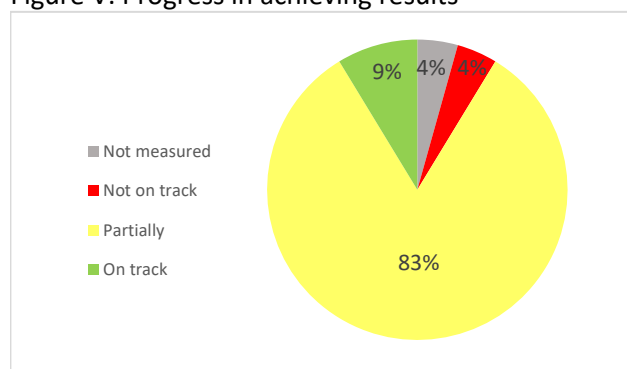
Besides limitations on reporting on results, the analysis focused on the extent to which evaluations faced limitations to engage with marginalized women. 83% of evaluations did not mention any limitations, which could be explained by different reasons: first, programme beneficiaries were not considered as stakeholders to participate in the evaluation; second, limitations imposed by the COVID-19 pandemic made it impossible to conduct field visits and obtain direct input from the most vulnerable groups; and third, the fact that this limitation is not mentioned in the evaluation reports may also be an indicator of the lack of participatory and intersectional approach.

Evaluations mentioning limitations on this issue, as well as extra measures undertaken, reported that attention was paid to the language barrier, to the participation of beneficiaries with trauma or to address high expectations among beneficiaries (refugees) that interviews would lead to resettlement. In connection to this issue, it is necessary to clarify that some of these limitations did not relate exclusively to marginalized women but to a broader category of beneficiaries.

2. Evaluation results

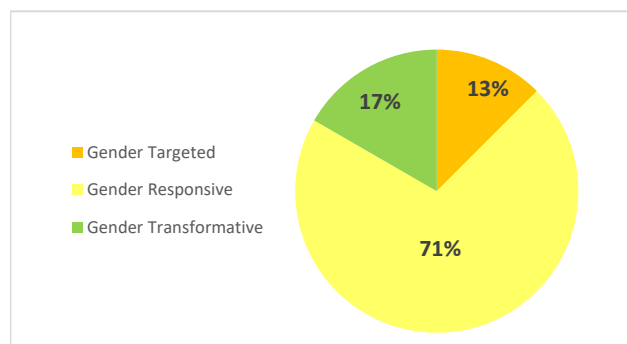
The second part of the analysis examines two issues: the achievement of results and the application of the Gender Results Effectiveness Scale (GRES) to the gender results.

Figure V. Progress in achieving results



A large majority of evaluations (83%) concluded that interventions results were partially achieved, while 9% showed that results were on track. In only one evaluation results were reported to be not on track on the grounds of the absence of a clear Theory of Change, while in another evaluation this analysis was not possible due to travel restrictions and lack of information available at the HQ level, making it impossible to provide a reliable examination of the progress made.

Figure VI. Gender results according to the GRES



The Gender Results Effectiveness Scale (GRES) was created to classify gender results according to five categories: Gender Negative, Gender Blind, Gender Targeted, Gender Responsive, and Gender Transformative, and is being increasingly used to determine the nature of gender related interventions. Based on the reviewed evaluations, gender results of the evaluations have been assigned a GRES category, particularly in the main findings and conclusions about the transformative character of programmes and policies. As Figure VI illustrates, most gender results (71%) are considered gender responsive, 17% gender targeted and 13% gender transformative.

Gender transformative interventions include work with non-traditional partners, critical awareness of gender roles and norms, or behavioral change at the community level, to mention just a few. In addition to these examples, specific attention is paid to the need to promote women’s rights and to challenge the underlying causes of gender inequality over the longer term to ensure the transformative potential of interventions.

EVALUATIONS REVIEWED

OIOS	Evaluation of Women, Peace and Security in field-based missions: Elections and Political Transitions
WIPO	Make Gender Equality a Habit. Evaluation-Audit Report of WIPO’s Policy on Gender Equality
OCHA	Inter-agency evaluation on gender equality and the empowerment of women and girls
UNESCO	From Ambition to Action: Evaluation of the UNESCO Global Priority Gender Equality
UNV	Evaluation of United Nations Volunteers’ Support to UN Peacebuilding Fund’s Gender Promotion Initiative (GPI)
WFP	Evaluation of the Gender Policy (2015-2020)
IOM	Final Evaluation for the Project: “Psychosocial and Medical Assistance to Urban Refugee Survivors and those at Risk of Sexual and Gender Based Violence (SGBV) in Kampala, Uganda”
UNECE	Programme level evaluation: Gender Mainstreaming in UNECE
ILO	Independent Final evaluation of “Strengthening gender monitoring and evaluation in rural employment in the Middle East and North Africa” project
DPI	Gender evaluation of the work of the Department of Global Communications Final Report
FAO	Evaluation of FAO’s Work on Gender
UNICEF	The UNICEF Multi-Country Gender-based Violence in Emergencies Evaluation Final Synthesis Report
UNFPA	Evaluation of UNFPA support to the prevention of, response to and elimination of gender -based violence and harmful practices (2012-2017)
JSDG Fund	Joint Programme on Rural Women’s Economic Empowerment (JP RWEE) in Ethiopia, End Evaluation Report
UNESCAP	Evaluation of Development Account Project (Code 415AO) Strengthened capacity of small island developing States in the Asia-Pacific region with regard to the valuation of environmental capital and the economic cost of gender inequality
OHCHR	Evaluation of the OHCHR Regional Gender Advisors Structure. Final Evaluation Report
UNECLAC	Final assessment report. Evaluation of the Strategy for mainstreaming gender at the Economic Commission for Latin America and the Caribbean (ECLAC), 2013–2017
IAEA	Evaluation of Gender Mainstreaming in Agency Programmes and Operations
UNAIDS	Evaluation of the UNAIDS Secretariat Gender Action Plan 2018-2023 Report and Annexes
UNCDF	Evaluation of UNCDF’s Strategic Framework 2018-2021
UNFPA	Evaluation of UNFPA support to gender equality and women’s empowerment (2012-2020)
UN-Habitat	Evaluation of UN-Habitat’s Policy and Plan for Gender Equality and Empowerment of Women in Urban Development and Human Settlements: 2014-2019
UNICEF	Realizing potential: evaluation of UNICEF’s Gender Action Plans
UNV	Evaluation of UNV Gender Equality and Women’s Empowerment for Organisational and Programmatic Effectiveness