



UNEG

United Nations Evaluation Group

Working Paper

UN-SWAP Evaluation Performance Indicator

2016 Reporting Cycle Results

This document is issued by the UNEG Working Group on Gender Equality and Human Rights and was drafted by Sabrina Evangelista, Evaluation Specialist and UN-SWAP Evaluation Performance Indicator Focal Point, UN Women with the support of Daniele Elizaire.

March 2017

FOREWORD

In 2016, the UNEG Norms and Standards were updated. The adoption of the stand-alone Norm 8 on Human Rights and Gender Equality is paramount for ensuring it is institutionalized across the UN system, including through the United Nations System-wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP).

In this context, UN-SWAP has continued to catalyze real change within UN entities practices with respect to gender equality. This report demonstrates the commitment and concrete actions taken by UN entities to enhance integration of gender equality in evaluation systems and processes.

To enhance UN-SWAP even further, UNEG undertook an Independent Review of UN-SWAP Reporting on the Evaluation Indicator that identified the key challenges faced by UNEG members in integrating gender-equality in evaluation as well as many good practices. As 2017 marks the final year of the first phase of UN-SWAP, the findings of this report are important for taking stock of where we are at and informing efforts to strengthen gender equality in evaluation, so that we can ensure UNEG members meet the requirements in compliance with the deadline, and strive to exceed requirements during the second phase of UN-SWAP – to commence in 2018.

I would like to congratulate all UNEG members that have taken concrete steps for improving integration of gender equality and human rights in their evaluation work – let's continue the trend!

Marco Segone
UNEG Chair

Table of Contents

FOREWORD	2
Acronyms	4
Executive Summary	5
I. Background	6
United Nations System Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP)	6
UN-SWAP Evaluation Performance Indicator (EPI).....	7
II. UN-SWAP EPI 2016 Reporting Cycle	8
Methodology.....	8
Evaluation Performance Indicator Results – 2016 Reporting Cycle	9
Finding 1: The majority (74%, N=32/43) of reporting entities are following the UNEG endorsed process for reporting.	9
Finding 2: Over half (56%, N=18/32) of entities that used the UNEG Scorecard sought an external perspective, which an Independent Review managed by UNEG has shown to ensure a more systematic application of UNEG Guidance on HR & GE.....	11
Finding 3: Over half (57%, N = 18/32) of scorecard users at least meet requirements, however, 44% (N=8/18) of these reports were based on an internal review.	13
Finding 4: Overall, the evaluation reports are just barely “satisfactorily integrating” the 4 UN-SWAP assessment criteria in evaluation reports; the weakest area of evaluation reports assessed continues to be gender-responsive methods.....	14
Finding 5: The majority (69%, N=22/32) of entities reported either planned or completed actions to institutionalize gender equality in evaluation systems.	16
Finding 6: The majority (81%, N=25/31) of UNEG Scorecard users have seen improvements in their score since 2014.....	17
Conclusion.....	17
III. Way forward	18
Annex 1: Overall 2016 UN-SWAP EPI Trends Reported to ECOSOC	19
Annex 2: Overall Rating by Type of Entity Reported to ECOSOC	19
Annex 3: UN Entities that used the UNEG Scorecard, 2016	20
Annex 4: Ratings for those entities that used the UNEG Scorecard, 2014-2016	23

Acronyms

CEB	Chief Executive Board for Coordination
ECOSOC	Economic and Social Council
EPI	Evaluation Performance Indicator
HR & GE	Human Rights and Gender Equality
PLE	Peer Learning Exchange
UNEG	United Nations Evaluation Group
UN-SWAP	United Nations System-Wide Action Plan on Gender Equality and the Empowerment of Women

Executive Summary

This document has been prepared to inform UNEG members on the 2016 reporting cycle results of the United Nations System-Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP) Evaluation Performance Indicator (EPI). The document provides a brief background of the Chief Executive Board for Coordination (CEB) endorsed UN-SWAP and UNEG support to UN entities to comply with the annual reporting against its EPI.

The key results from 2016:

Finding 1: The majority (74%, N=32/43) of reporting entities are following the UNEG endorsed process for reporting.

Finding 2: Over half (56%, N=18/32) of entities that used the UNEG Scorecard sought an external perspective, which an Independent Review managed by UNEG has shown to ensure a more systematic application of UNEG Guidance on HR & GE.

Finding 3: Over half (57%, N = 18/32) of scorecard users at least meet requirements, however, 44% (N=8/18) of these reports were based on an internal review.

Finding 4: Overall, the evaluation reports are just barely “satisfactorily integrating” the 4 UN-SWAP assessment criteria in evaluation reports; the weakest area of evaluation reports assessed continues to be gender-responsive methods.

Finding 5: The majority (69%, N=22/32) of entities reported either planned or completed actions to institutionalize gender equality in evaluation systems.

Finding 6: The majority (81%, N=25/31) of UNEG Scorecard users have seen improvements in their score since 2014.

The report concludes that UNEG members have made commendable progress in integrating gender equality in evaluation since the introduction of the UN-SWAP – overall scores have increased and more entities are seeking external perspectives on their progress. The next step is to facilitate a common understanding on the application of a gender-responsive evaluation approach amongst UNEG members. The way forward section identifies how the UN-SWAP 2.0, which will come into effect in 2018, may impact evaluation offices, and calls upon UNEG members to make dedicated efforts to evaluate progress of gender mainstreaming at an institutional level.

I. Background

United Nations System Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP)

The UN-SWAP constitutes the first accountability framework for gender mainstreaming in the UN system. Economic and Social Council (ECOSOC) Resolution E/RES/2014/12 requests the United Nations system, including its agencies, funds and programmes, within their respective organizational mandates, to continue working collaboratively to enhance and accelerate gender mainstreaming within the United Nations system, including by fully implementing the UN-SWAP.¹

The UN-SWAP is composed of 15 performance indicators for tracking 6 main elements on gender mainstreaming: accountability, results based management, oversight, human and financial resources, capacity, and knowledge exchange and networking. All UN entities are to self-assess and report on progress. UN entities are expected to meet all UN-SWAP performance standards by 2017². During 2016, UN-SWAP 2.0 has been finalized, which will result in a revised UN-SWAP framework to be rolled out in 2018.³

Reporting on the UN-SWAP commenced in 2013 (on the 2012 reporting cycle) and entities are expected to report on a yearly basis through the Report of the Secretary-General to ECOSOC on “Mainstreaming a gender perspective into all policies and programmes in the UN system”⁴. UN Women provides secretariat services and support to the UN system for reporting on the UN-SWAP. The Gender units of all UN entities play an internal coordinating role, ensuring that progress is accurately reported and that plans of action are developed. Accountability rests, as noted in the CEB policy, with senior managers of the different UN-SWAP reporting entities.

The UN-SWAP framework is accompanied by a set of Technical Notes for each Performance Indicator that provide guidance on how to complete the assessment for each of the 15 Performance Indicators. While the UN-SWAP Performance Indicators approved by the CEB are established, the Technical Notes are considered live documents that can be enhanced.

¹ UN-SWAP Framework was developed by UN Women in 2011/2012 in response to the CEB endorsed UN system-wide Policy on Gender Equality and the Empowerment of Women (CEB/2006/2), which was established based on the ECOSOC Agreed Conclusions 1997/2. On 13 April 2012, the CEB endorsed the UN-SWAP for application throughout the UN system. In its resolution E/RES/2012/24 of 27 July 2012, the ECOSOC welcomed the UN-SWAP and called upon the UN system to actively engage in its roll-out and report on the implementation of the resolution at its substantive session in 2013.

² There is an extended timeframe to 2019 for those entities with a mainly technical focus.

³ UN-SWAP/2 framework has been endorsed by UN system Gender Focal Points, and now the Technical Notes are under development.

⁴ UN-SWAP Evaluation Performance Indicator 2015 Reporting Cycle results are accessible online: <http://www.uneval.org/document/download/2536>.

UN-SWAP Evaluation Performance Indicator (EPI)

The oversight element of the UN-SWAP is composed of three performance indicators, including one dedicated to evaluation that is linked to meeting the gender-related UNEG Norms & Standards⁵ and demonstrating effective use of the UNEG guidance on integrating human rights and gender equality (HR & GE) in evaluation⁶.

The reporting categories for the Evaluation Performance Indicator (EPI) are as follows⁷:

Table 1. Evaluation Performance Indicator				
Not Applicable	Missing	Approaches	Meets	Exceeds
5a. Performance indicator is not relevant to a UN entity.	5b. None of the UNEG gender-related norms and standards are met.	5c. Meets some of the UNEG gender-related norms and standards.	5d. Meets the UNEG gender-related norms and standards.	5ei. Meets the UNEG gender-related norms and standards and 5eii. Demonstrates effective use of the UNEG guidance on evaluating from a human rights and gender equality perspective.

The UNEG Working Group on Gender Equality and Human Rights developed the Technical Note and Scorecard, which aims to support more systematic and harmonized reporting through a common tool that allows for improved comparability across UN entities. UNEG Heads endorsed the Technical Note and Scorecard in August 2014 (available on the UNEG website: <http://www.uneval.org/document/detail/1452>) and in 2016, a Frequently Asked Questions was produced by the working group to support reporting (<http://www.unevaluation.org/document/download/2631>). The unit of analysis selected as most feasible was the evaluation report. Thus, the Evaluation Performance Indicator should be solely based on an assessment of evaluation reports completed in the reporting year. The Technical Note specifies the below criteria for the assessment of integration of gender equality in the evaluation reports:

⁵ United Nations Evaluation Group, Norms and Standards, updated in 2016: <http://www.unevaluation.org/document/detail/1914>

⁶ United Nations Evaluation Group, Guidance on Integrating Human Rights and Gender Equality in Evaluation, 2014; accessible online: <http://unevaluation.org/document/detail/1616>

⁷ Please note that in 2016 revisions to the evaluation indicator were endorsed by UNEG members, which will come into effect in 2018.

Table 2. UN-SWAP Evaluation Performance Indicator criteria for assessing evaluation reports

Criteria 1	GEEW is integrated in the evaluation scope of analysis and evaluation indicators are designed in a way that ensures GEEW-related data will be collected.
Criteria 2	GEEW is integrated in evaluation criteria and evaluation questions specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.
Criteria 3	A gender-responsive methodology, methods and tools and data analysis techniques are selected.
Criteria 4	The evaluation findings, conclusions and recommendations reflect a gender analysis.

It is recommended that evaluation units conduct an external review (hiring a consultant or firm). However, as this requires financial resources, the UNEG Working Group on Gender Equality and Human Rights has promoted a Peer Learning Exchange (PLE) process for the past two reporting cycles, which has proven to be successful. The PLE involves a minimum of two entities reviewing each other's evaluation reports (or a sample of) against the UNEG endorsed EPI Evaluation Scorecard, comparing results of the entities self-assessed EPI report, and providing feedback. One key purpose of the PLE is to strengthen and provide quality control for annual reporting on UN SWAP evaluation performance indicator and the corresponding remedial plan of action. At a minimum, UN entities are highly encouraged to submit the UNEG endorsed EPI Evaluation Scorecard even when based on internal/self-assessment to allow for comparability. The below analysis is disaggregated according to the aforementioned categories: external, PLE or internal assessment. Through the UNEG Working Group on Gender Equality and Human Rights, UN Women Independent Evaluation Office⁸ provides training webinars on the UN-SWAP EPI reporting process and Scorecard, help desk support, and drafts the trend report.

II. UN-SWAP EPI 2016 Reporting Cycle

Methodology

The following results are based on the data that was input for the UN-SWAP EPI by the respective entity UN-SWAP focal point into the online UN-SWAP reporting database⁹, which is administered by UN Women. The UN Women EPI focal point accessed the information from the database, accompanying attachments (e.g. the meta-analysis, scorecard, etc.) and narrative report and liaised with the respective EPI focal point to ensure the information was accurate. It is important to keep in mind that the reporting practices of UNEG members varies greatly, limiting the ability to generalize results across entities or compare with previous reporting cycles. Even when entities follow the UNEG reporting process or conduct an external review, there is much variation according to interpretation of the criteria, and the number (between 1-61 reports) and type (produced by central offices vs. decentralized offices) of reports reviewed. Therefore, the interpretation of results should take this into consideration.

⁸ The current focal point is Sabrina Evangelista: Sabrina.evangelista@unwomen.org

⁹ <https://unswap.unwomen.org>

Evaluation Performance Indicator Results – 2016 Reporting Cycle

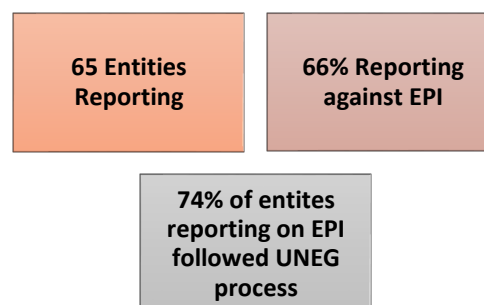
For the 2016 UN-SWAP reporting cycle 65 UN entities submitted¹⁰ a report. Of these, 22 (N=22/65, 34%) entities reported that the EPI was not applicable, which is recommended when the entity does not have an evaluation unit or carry out evaluation as defined by UNEG.¹¹ The majority (N=15/22, 68%) of entities that reported “not applicable” were from the Secretariat.

UN Women reports the aggregate results (including those that did not follow UNEG endorsed reporting process) to the UN General Assembly (see Annex 1 & 2 for details). As noted, these ratings are not comparable because entities that do not follow the UNEG endorsed reporting process base their assessment on miscellaneous factors (such as gender balance of evaluation teams), whereas the UNEG endorsed Technical Note requires that the rating be based on an assessment of evaluation reports. Therefore, to enhance comparability, findings 2-4 are based specifically on those entities that followed the UNEG endorsed EPI reporting process.¹²

Finding 1: The majority (74%, N=32/43) of reporting entities are following the UNEG endorsed process for reporting.

UNEG promotes the use of the Technical Note and accompanying Scorecard to have a systematic unit of comparison (the evaluation report), criteria against which to assess and scoring categories corresponding to the UN-SWAP categories. Those evaluation units that did not use the UNEG scorecard¹³ based their rating on other factors, such as hiring processes (the extent to which females are hired to conduct evaluations) or

Figure 1. 2016 At a Glance



¹⁰ The following entities did not report against the UN-SWAP indicators for the 2016 reporting cycle: United Nations Interregional Crime and Justice Research Institute, the United Nations Institute for Disarmament Research, the United Nations Research Institute for Social Development; the Office of the Special Representative of the Secretary-General on Sexual Violence in Conflict; and the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization.

¹¹ The following UN-SWAP reporting entities reported that the Evaluation Indicator was not applicable in 2016: CAAC, DGAM, DM, DPA, ICAO, ITU, OAJ, ODA, OHRLS, OLA, Ombudsman, OSAA, UNFCCC, UNISDR, United Nations Global Compact, UNOG, UNON, UNOPS, UNU, UNV, UNWTO, and WMO.

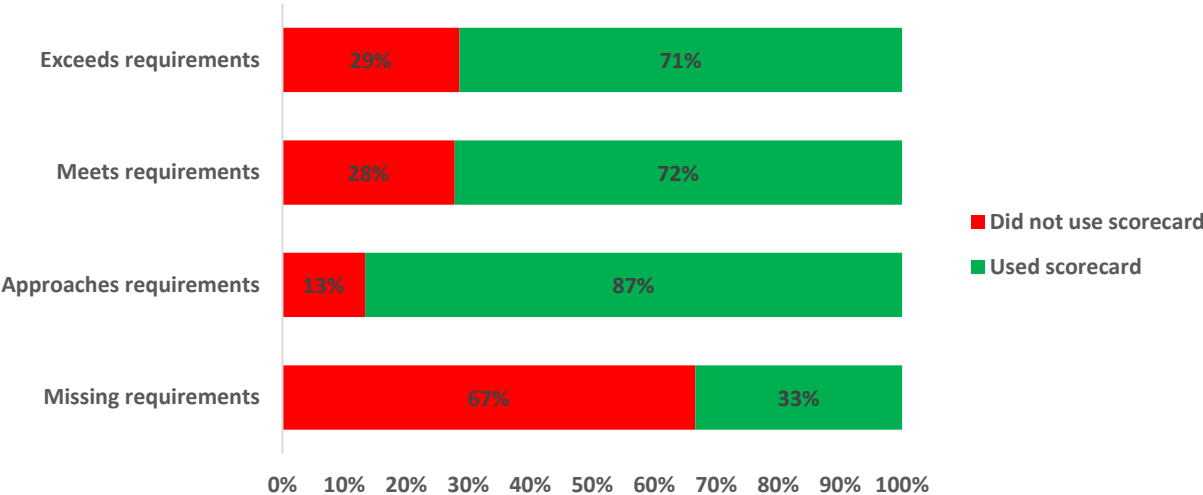
¹² Please note that although UNOV apparently reports in aggregated form with UNODC for all UN-SWAP indicators, the EPI reporting has only been completed for UNODC. UNOV does not have an evaluation unit and is not a UNEG member. Therefore, UNOV did not report on the EPI and was not included in the detailed, disaggregated analysis below. On the other hand, the GEF is not an official UN-SWAP reporting entity, but is a UNEG member and reports on an annual basis, thus it is included in the below analysis but not in the aggregate score reported to ECOSOC.

¹³ Entities that did not follow the UNEG endorsed process include: DESA, DFS, DPKO, ECA, ECE, IAEA, IMO, ITC-ILO, UNHCR (although UNHCR did participate in PLE a scorecard was not submitted), UNSSC, and UPU.

systems in place to support gender equality in evaluation (such as an organizational policy on gender equality). The UNEG scorecard allows for a minimum level of coherence in reporting.

In 2016, the majority of reporting entities used the UNEG Scorecard and Technical Note: 74% (N=32/43).¹⁴

Figure 2. UNEG scorecard usage as per cent of total rating category (N=43)



¹⁴ 2016 data includes GEF.

Finding 2: Over half (56%, N=18/32) of entities that used the UNEG Scorecard sought an external perspective, which an Independent Review managed by UNEG has shown to ensure a more systematic application of UNEG Guidance on HR & GE.

In 2016, **10 (31%) submissions were based on an external review by external companies/consultants**¹⁵: FAO, IFAD, ILO, UNDP, UNEP, UNFPA, UNICEF, UNODC, UN Women, and WFP; **8 (25%) entities participated in the PLE**: UN ESCAP, OHCHR, UNCDF, UNCTAD, UN-Habitat, UNIDO, UNRWA, IOM, [and UNHCR¹⁶]; and **14 (44%) entities based their assessment on an internal review**: DPI, DSS, ECLAC, ESCWA, GEF, ITC, OCHA, PBSO, UNAIDS, UNESCO, UNITAR, WHO, WIPO, and OIOS. See Table 3 for details.

The results of external assessments are more objective, allowing for more comparability across entities.¹⁷ The Independent Review of UN-SWAP Evaluation Indicator Reporting¹⁸ [The Independent Review], which reviewed 46 evaluation reports completed in 2015 from 23 UNEG member entities, confirmed that **externally and PLE reviewed reports were more systematic in applying UNEG Guidance on HR & GE.**

The Independent Review found that **there were significant discrepancies between ratings included in the original scorecard and the independent review. For example, 81 per cent of the reports originally rated as “exceeding requirements” were re-classified as “meeting”, “approaching” or “missing” requirements**¹⁹. In fact, three reports rated as “exceeding requirements” by the original scorecard assessment, were rated as “missing requirements” by the Independent Review and did not even meet the

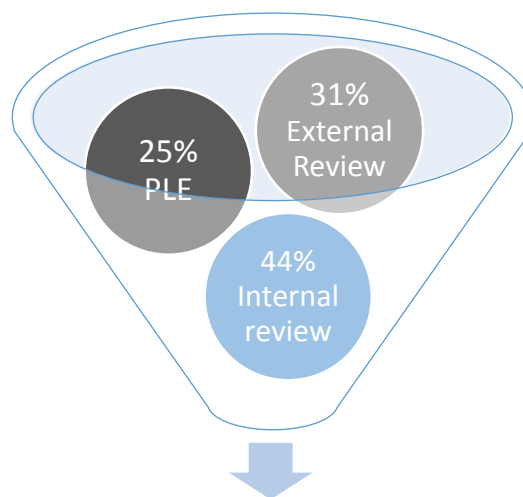


Figure 3. 2016 Reporting Cycle

¹⁵ Both ILO and UNICEF submissions were based on review of evaluation reports completed in 2015 due to the misalignment of reporting processes. UNEG Heads endorsed on a non-objection basis that an external review could be based on evaluation reports completed in the previous year. Efforts to align meta-evaluation processes with the UN-SWAP reporting cycle are encouraged.

¹⁶ Although UNHCR participated in the Peer Learning Exchange, it did not submit a scorecard for its review and thus is not included in the overall statistics presented here.

¹⁷ As the overall UN-SWAP was built as a self-assessment tool, it is important to note that the Evaluation Indicator is the only category that advocates for an external assessment because external assessments are more objective than internal/self-assessments. All other categories of the UN-SWAP are assessed internally.

¹⁸ UNEG Working Group on Gender Equality and Human Rights, Independent Review of UN-SWAP Evaluation Performance Indicator Reporting, 2016; accessible online at: <http://www.uneval.org/document/download/2634>.

¹⁹ Out of the 17 reports that were re-classified by the independent assessment, 9 were re-classified as ‘meeting requirements’, 5 as ‘approaching requirements’ and 3 as ‘missing requirements’.

general UNEG evaluation standards. Out of the 17 reports that were re-classified to a lower level by the independent assessors, 76 per cent of these were internally reviewed reports.

The Independent Review found that differences between the original UN-SWAP EPI scorecard assessments (conducted by UN entities, external companies/consultants and the PLE) and the independent assessment reflected inconsistent understandings of what it means to “exceed requirements”. These discrepancies also demonstrated how difficult it is to compare and assess adherence to UN-SWAP criteria across UN entities based on the internal/self-assessment rating.

The good news is that in 2016 more entities sought an external perspective against the UN-SWAP EPI criteria: 31% (N=10/32) of submissions were based on external review²⁰ and 25% (N=8/32)²¹ participated in the Peer Learning Exchange²² for the UN-SWAP EPI. When we combine the external and PLE, this brings us to over half (56%, N=18/32) of reporting entities that used the UNEG scorecard reporting with an external perspective. Although the number of entities conducting external reviews did not increase from 2015, the number of participants in the PLE doubled.

Table 3. Disaggregated results for UN entities using the UNEG Scorecard, 2016 (N=32)						
# Reports →	1-2 Reports	3-5 Reports	6-10 Reports	11-15 Reports	16-20 Reports	21 or more reports
Rating ↓						
Exceeds	OHCHR^ UNCDF^	PBSO ITC	ESCWA			
Meets		UNAIDS UNCTAD^ WIPO ECLAC WHO	UN-Habitat^ UNIDO^ ESCAP^	UNFPA*	IFAD* UNESCO	WFP* UN Women*
Approaches	DPI OCHA	UNITAR UNRWA^	IOM^ GEF OIOS		UNODC*	UNEP* FAO* UNDP* UNICEF* ILO*
Missing	DSS					
^Participated in PLE *External Review						

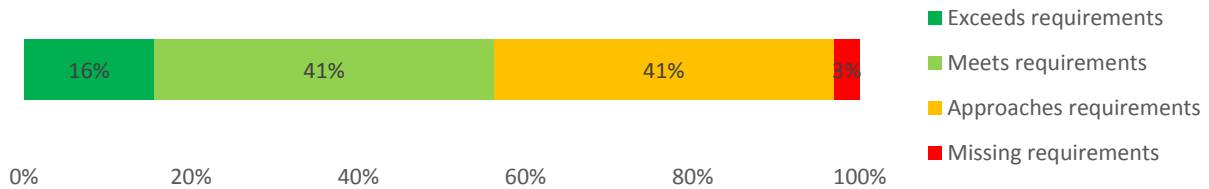
²⁰ Entities which conducted an external review include: FAO, IFAD, ILO, UNDP, UNEP, UNFPA, UNICEF, UNODC, UNW, and WFP.

²¹ UNODC conducted both an external review and a Peer Learning Exchange. UNHCR also participated in the PLE, however, did not submit a scorecard and is thus not included in the overall statistics.

²² Entities that engaged in a Peer Learning Exchange include: IOM, OHCHR, UNCDF UNCTAD, UN-Habitat, UNIDO, UNODC, and UNRWA [and UNHCR].

Finding 3: Over half (57%, N = 18/32) of scorecard users at least meet requirements, however, 44% (N=8/18) of these reports were based on an internal review.

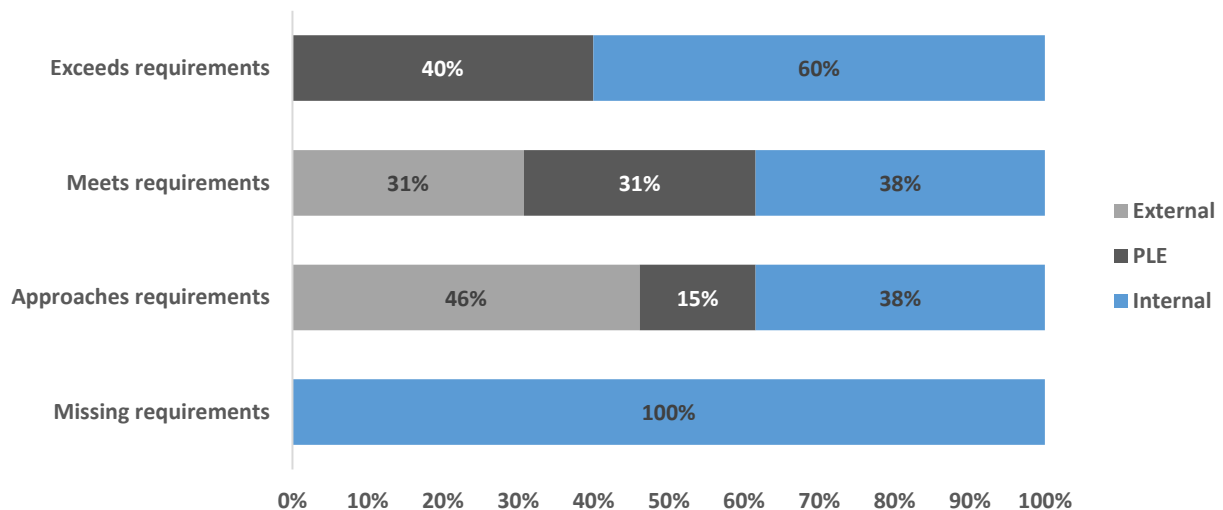
Figure 4. Aggregate Evaluation Performance Indicator Rating Among Scorecard Users (N=32)



With 41% (N=13/32) of entities “meeting requirements” and 16% (N=5/32) “exceeding requirements”, more than half of scorecard users (57% together) have reached this benchmark for gender responsive evaluation.

However, as can be seen in Figure 5, none of the 5 entities that scored “exceeds requirements” hired an external reviewer – although two (OHCHR and UNCDF) participated in the PLE; and 38% (N=5/13) of entities that reported “meets requirements” based their rating on an internal review.

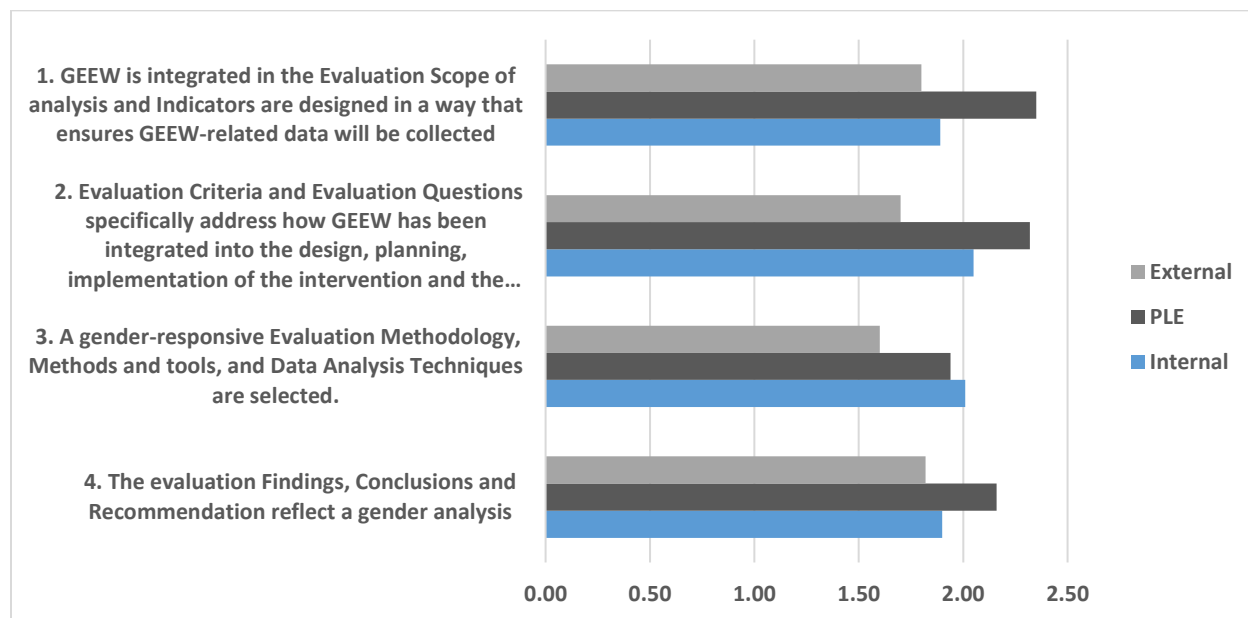
Figure 5. Type of review as percent of total reporting category (N=32)



Finding 4: Overall, the evaluation reports are just barely “satisfactorily integrating” the 4 UN-SWAP assessment criteria in evaluation reports; the weakest area of evaluation reports assessed continues to be gender-responsive methods.

In 2016, 449 evaluation reports were scored against the EPI criteria. The average score by UN-SWAP criteria was calculated and disaggregated by type of review. Overall, the reports are just barely “satisfactorily integrating” the 4 criteria in evaluation reports (a score of 2 indicates satisfactorily integrating). The weakest area of evaluation reports assessed continues to be gender-responsive methods (as was reported last year) and corroborated by the Independent Review.

Figure 6. Average score by criteria and type of review for UNEG scorecard users (N=32)



The Independent Review of UN-SWAP EPI reporting identified the key challenges that UNEG members experience in meeting the requirements across the criteria, as outlined in Table 4. These findings should be used for strengthening efforts. UNEG is also producing a good practice guide based on the report’s findings, which will be issued in April 2017.

Table 4. Findings from Independent Review²³	
Criteria	Findings
Criteria 1: Scope and Indicators	<p>There is an inconsistent understanding about how to integrate GEEW in evaluation scope of analysis</p> <ul style="list-style-type: none"> • Very few reports included an analysis of the extent to which GEEW was addressed in the project results framework or theory of change.

²³ UNEG Working Group on Gender Equality and Human Rights, Independent Review of UN-SWAP Evaluation Performance Indicator Reporting, 2016; accessible online at: <http://www.uneval.org/document/download/2634>.

	<ul style="list-style-type: none"> • Reports did not consistently assess the extent to which gender issues and underlying causes of inequality, were included in situational/contextual analyses. • Most reports considered outputs or outcomes only where gender was explicitly addressed. For interventions where GEEW was not an explicit focus and sex-disaggregated data was not readily available, evaluations tended not to draw upon established theory to discuss the likely gender considerations or how GEEW could have been better integrated. • A very limited number of evaluations discussed GE & HR standards and the extent to which the design of interventions supported implementation of relevant standards such as CEDAW and national policies and strategies related to GEEW. • A very limited number of evaluations included an analysis of relevant organizational or UN system-wide objectives or strategies on GEEW and their relevance to the object of the evaluation. • Inclusion of sex-disaggregated data was missing from most evaluation reports. • Where sex-disaggregated data was available, evaluations often did not go beyond number counting to discuss the extent of women’s and girl’s participation and how interventions addressed (or not) their needs and underlying causes of gender inequalities. • Most of the GEEW-related indicators included in the Evaluation Matrices were specific to gender-related questions and there was inadequate mainstreaming of GEEW across other indicator areas in most reports. • Where GEEW was not an explicit focus, some evaluations looked more broadly at how empowerment approaches were applied, how human rights was mainstreamed or how interventions contributed to greater accountability towards specific populations (for example humanitarian response-related interventions) or how they addressed issues of social cohesion.
<p>Criteria 2: Criteria and Questions</p>	<p>There is a general understanding about how to integrate GEEW in evaluation criteria and questions although approaches in applying the criterion are somewhat inconsistent and uneven:</p> <ul style="list-style-type: none"> • For reports internally assessed, reports were often rated as ‘exceeds’ in instances where GEEW was included in one question and not any of the criteria; whereas for the independently reviewed reports, a rating of ‘exceeds’ was only given in instances where GEEW was integrated in multiple criteria and questions. • The review also found that including GEEW as a stand-alone criterion generally ensured that it was addressed in the report but also resulted in a lower level of mainstreaming across the other criteria. Very few reports achieved systematic integration of GEEW across all criteria.
<p>Criteria 3: Methods</p>	<p>Gender responsive methods are the weakest aspect of the overall body of reports, and are insufficient to meet the required UN-SWAP standard. More than half of the reports independently assessed failed to meet or exceed requirements:</p> <ul style="list-style-type: none"> • Many reports included a general statement that gender-responsive methods were used but most reports lacked sufficient detail about how gender equality was incorporated into the evaluation design and approach and, more specifically, how analytical methods were used to draw out gender implications. • For the majority of reports, there was limited detail on how gender was integrated within data collection tools and analysis methods. • Participatory approaches and the inclusion of women and/or girls was often mentioned within evaluation processes, however, most reports did not include

	comprehensive stakeholder analysis whereby barriers to participation, including potential unequal power relations, were identified and mitigation measures to maximize inclusion were discussed.
Criteria 4: Findings, Conclusions and Recommendations	<p>Approaches to mainstreaming GEEW across the findings, conclusions and recommendations varied and a systematic approach to weaving gender more comprehensively across all three was often missing:</p> <ul style="list-style-type: none"> • Reports that effectively integrated GEEW into the evaluation scope of analysis and across multiple criteria and questions, consistently mainstreamed GEEW across findings, conclusions and recommendations. Where reports included a stand-alone section on GEEW, this usually resulted in a specific gender-related finding but not always a reflection of GEEW in the conclusions and recommendations. • Evaluation reports also often concluded that where data was limited, there was insufficient information to conduct a gender analysis, thus passing responsibility back to the entity. Whilst most evaluation reports rely on primary monitoring data, without this, evaluators can still look at the wider body of theory and sources of secondary data in assessing how a particular intervention was designed (or not designed) to address gender.

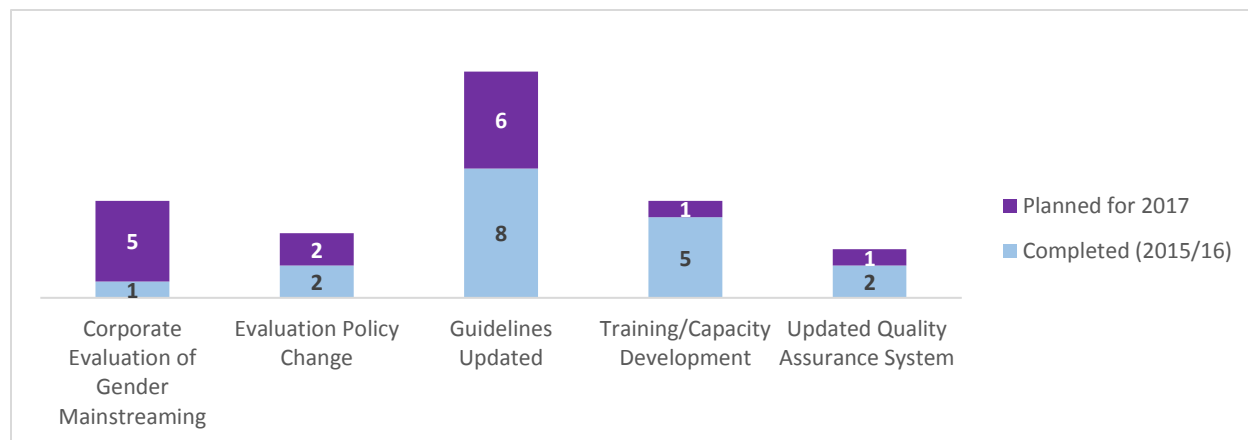
Finding 5: The majority (69%, N=22/32) of entities reported either planned or completed actions to institutionalize gender equality in evaluation systems.

All self-reported narrative comments submitted as part of the UN-SWAP EPI submission were reviewed in order to identify common actions taken by reporting entities to institutionalize gender equality in evaluation systems.²⁴ Five unique types of progress have been made: 1) changes to evaluation policy; 2) updates to evaluation guidelines; 3) updated quality assurance systems; 4) training/capacity development; and 5) institutional evaluations of gender mainstreaming. Twenty-two entities reported 33 actions to institutionalize gender equality in evaluation systems, whether completed in 2015/2016 or planned for 2017– some entities reporting more than one concrete action over the past two years.²⁵ Of those actions, 54% (N=18/33) have achieved completion in 2016. Updating evaluation guidelines to incorporate gender equality was the most frequently occurring category representing 42% (N=14/33) of total actions.

²⁴ Not all entities report on actions taken, thus this may not be comprehensive of all actions taken by all entities, but rather provides a useful overview.

²⁵ The following entities referred to more than one type of action to institutionalize gender quality in evaluation systems: IFAD, ILO, IOM, UNDP, UNICEF, UNV, and WFP.

Figure 7. Number of actions by type of action and completion status (N= 33)



Finding 6: The majority (81%, N=25/31) of UNEG Scorecard users have seen improvements in their score since 2014.

The overall numerical score reported between 2014 and 2016 were compared to identify trends (see Annex 4), as the Technical Note was revised for implementation in 2014 (thus, previous scores are not comparable). Of the 31 reporting entities that used the UNEG scorecard during 2014 or 2015²⁶, 81% have reported an increase in their overall score in 2016 (including only one that had the same score). However, less than half of these scores were based on external assessments. Nevertheless, the trend is positive for integration of gender equality in evaluation.

Conclusion

After five years of UN-SWAP reporting²⁷, progress integrating gender equality in evaluation systems across the UN system is commendable. This progress is thanks to the commitment and dedication of UNEG members in identifying ways of applying the UNEG Norms and Standards in their unit's work. Additionally, more evaluation offices are seeking external perspectives in their UN-SWAP assessment adding to the robustness of the findings. As the Independent Review of UN-SWAP EPI reporting illustrated, there is still room for improvement in ensuring a common understanding about what it means to integrate gender equality in evaluation and thus to meet the requirements for UN-SWAP.

²⁶ Six of the entities used the scorecard in either 2014 or 2015 but not in both years, while 2 entities started to use the scorecard only in 2016 and thus were not included in this analysis.

²⁷ Annex 4 and Finding 6 only go back to 2014 because the Technical Note and criteria were revised and implemented in 2014.

III. Way forward

The revised UN-SWAP 2.0 framework provides an opportunity for re-doubling efforts to enhance integration of gender equality not only at an institutional level but also in evaluation work. The new indicators introduced require reporting whether the “Entity has achieved or is on track to achieve the high-level result [as outlined in the Strategic Plan equivalent] on gender equality and the empowerment of women”²⁸, which should be aligned with the SDGs. Evaluation offices should be ready to provide support to their respective institution for measuring progress of the entity related to gender equality and the empowerment of women.

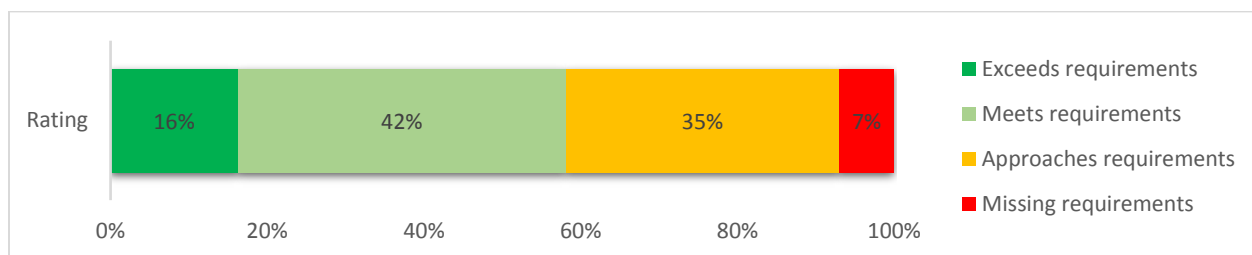
The UNEG endorsed revised EPI introduces the completion of an institutional evaluation of gender mainstreaming every 5-8 years to “exceed requirements”. This institutional evaluation of gender mainstreaming can provide the very data that the organization needs to report on the new UN-SWAP results indicators. The hope is that the new evaluation indicator will incentivize UNEG members to undertake these evaluations to not only address the institutional needs but also add to an enhanced and coherent understanding of progress mainstreaming gender across the UN system and the possible effects on development results. The UNEG Working Group will be continuing its efforts to support UNEG members by issuing in May 2017 a Guidance Document for evaluating gender mainstreaming; and in 2017 will undertake the revision of the Technical Note for the UN-SWAP reporting to enhance clarity on the criteria and how to “meet or exceed requirements”.

Over the past 5 years we have seen an increase in awareness and commitment to integrating gender equality in evaluation. As the trend reports and Independent Review have shown it is now time to enhance and build a common understanding about what it means to apply a gender-responsive approach in evaluation. There are many good practices that have been identified and the focus moving forward will be to codify and share these practices, which will increase awareness regarding how to integrate gender equality and contribute towards a common understanding. The UNEG Working Group will be issuing a good practice guide in April 2017 to facilitate this understanding and conducting a series of webinars to facilitate exchange.

²⁸ UN Women Coordination Division, UN-SWAP 2.0 Indicators, December 2016.

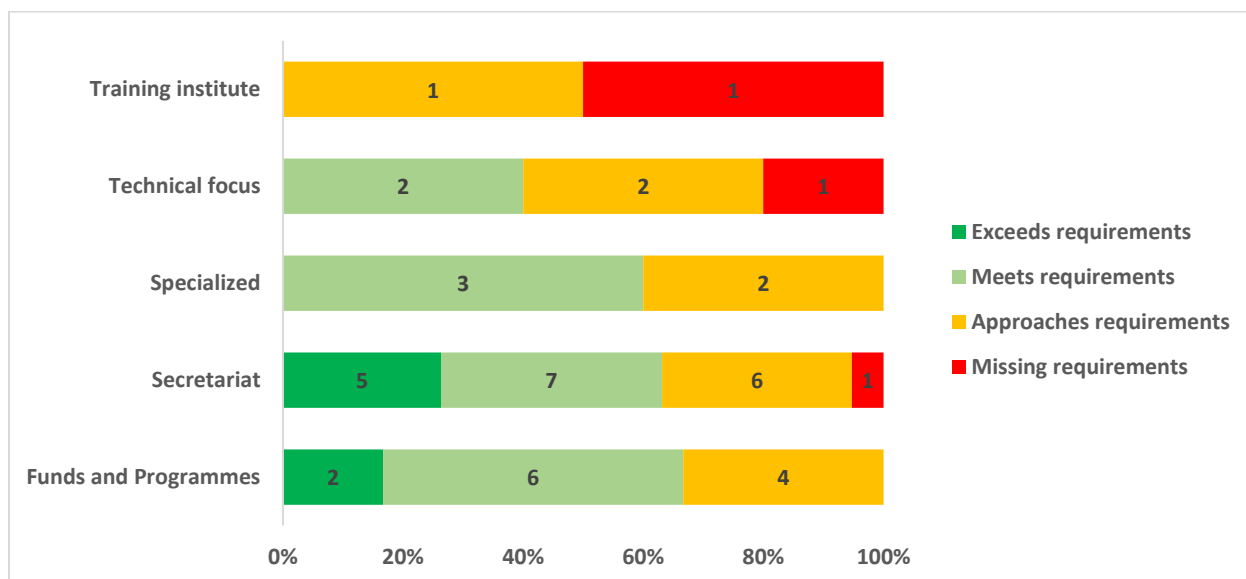
Annex 1: Overall 2016 UN-SWAP EPI Trends Reported to ECOSOC

(including entities that did and did not use UNEG scorecard, N=43)



Annex 2: Overall Rating by Type of Entity Reported to ECOSOC

(including entities that did and did not use UNEG scorecard, N=43)



Annex 3: UN Entities that used the UNEG Scorecard, 2016

Type of Review	Entity Type	Entity Short Name	Ratings	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Overall Score	# Reports Reviewed	Range of reports
External	Secretariat	UNODC	Approaches	1.63	1.63	1.37	1.89	6.53	19	equal to 16-20
External	Specialized	FAO	Approaches	1.21	1.39	1.87	1.64	6.12	28	equal to 21 or more
External	Funds and Programmes	IFAD	Meets	2.72	2.72	2.56	2.5	10.5	18	equal to 16-20
External	Funds and Programmes	UNDP	Approaches	1.28	1.36	1.33	1.47	5.44	36	equal to 21 or more
External	Secretariat	UNEP ²⁹	Approaches	1.06	1.09	0.70	0.80	3.65	34	equal to 21 or more
External	Funds and Programmes	UNFPA	Meets	2.85	2.15	1.77	2.85	9.62	13	equal to 11-15
External	Funds and Programmes	UNICEF	Approaches	1.7	1.7	1.6	1.4	6.3	61	equal to 21 or more
External	Funds and Programmes	UN Women	Meets	2.2	2.3	1.7	2.2	8.4	36	equal to 21 or more
External	Funds and Programmes	WFP	Meets	2.14	1.66	2.14	2.14	8.07	29	equal to 21 or more

²⁹ UNEP submission was late and thus not reflected accurately in the ECOSOC report.

External	Specialized	ILO	Approaches	1.2	1.1	1	1.3	4.6	42	equal to 21 or more
Internal	Secretariat	DPI	Approaches	2	2	2	1	7	1	equal to 1-2
Internal	Secretariat	DSS	Missing	0	1	1	1	3	1	equal to 1-2
Internal	Secretariat	ECLAC	Meets	2	2.6	2	2.4	9	5	equal to 3-5
Internal	Secretariat	ESCWA	Exceeds	2.17	3	3	3	11.17	6	equal to 6-10
Internal	Funds and Programmes	GEF	Approaches	0.71	1	1.29	1.43	4.43	7	equal to 6-10
Internal	Funds and Programmes	ITC	Exceeds	3	3	3	3	12	4	equal to 3-5
Internal	Secretariat	OCHA	Approaches	1.5	1.5	1	1	5	2	equal to 1-2
Internal	Secretariat	PBSO	Exceeds	3	3	2.67	3	11.67	3	equal to 3-5
Internal	Funds and Programmes	UNAIDS	Meets	2	2	2	2	8	3	equal to 3-5
Internal	Specialized	UNESCO	Meets	2.1	2	2.1	2.3	8.5	20	equal to 16-20
Internal	Training institute	UNITAR	Approaches	2.33	1.67	1.67	1.33	7	3	equal to 3-5
Internal	Specialized	WHO	Meets	1.8	1.8	2.2	1.8	7.6	5	equal to 3-5
Internal	Technical focus	WIPO	Meets	2	2	2	2	8	4	equal to 3-5
Internal	Secretariat	OIOS	Approaches	1.86	2.14	2.14	1.29	7.43	7	equal to 6-10
PLE	Funds and Programmes	IOM	Approaches	1.5	1.5	1.5	1.67	6.17	6	equal to 6-10

PLE	Secretariat	OHCHR	Exceeds	3	3	3	2	11	1	equal to 1-2
PLE	Funds and Programmes	UNCDF	Exceeds	3	3	3	2	11	2	equal to 1-2
PLE	Secretariat	UNCTAD	Meets	2	2	1.75	2	7.75	4	equal to 3-5
PLE	Secretariat	UN-Habitat	Meets	2.67	2.33	2	3	10	6	equal to 6-10
PLE	Technical focus	UNIDO	Meets	2.39	2.36	1.39	1.43	7.57	30	equal to 21 or more
PLE	Funds and Programmes	UNRWA	Approaches	1.6	1.6	1.4	2.2	6.8	5	equal to 3-5
PLE	Secretariat	ESCAP	Meets	2.63	2.75	1.5	3	9.88	8	Equal to 6-10

Annex 4: Ratings for those entities that used the UNEG Scorecard, 2014-2016³⁰

Entity short name	2014			2015			2016		
	Rating	Score	#reports	Rating	Score	#reports	Rating	Score	# reports
IFAD*	Meets	9.66	15	Meets	10.2	21	Meets	10.50	18
UNICEF*	Approaches	6	15	Approaches	5.74	69	Approaches	6.3	61
UNEP*	Missing	2.2	17	Missing	2.9	35	Approaches	3.65	34
ILO*	Did not use scorecard			Approaches	4.6	42	Approaches	4.6	42
UNDP*	Approaches	5.9	30	Approaches	6	25	Approaches	5.44	36
WFP*	Meets	7.26	19	Meets	8.19	21	Meets	8.07	29
UN Women*	Meets	7.57	22	Approaches	6.6	27	Meets	8.4	36
UNESCO	Approaches	3.53	19	Approaches	3.8	15	Meets	8.5	20
DPI	Meets	9	1	Approaches	6	2	Approaches	7	1
OHCHR^	Meets	7.66	3	Exceeds	10.66	3	Exceeds	11	1
FAO*	Meets	7.68	28	Meets	8.54	22	Approaches	6.12	28
IOM^	Approaches	4.73	11	Approaches	5.25	8	Approaches	6.17	6
OCHA	Meets	9	1	Meets	9	2	Approaches	5	2
ECLAC	Approaches	7.25	4	Meets	10.125	8	Meets	9	5
UNCTAD^	Approaches	6.71	7	Approaches	7.43	7	Meets	7.75	4
ESCWA	Meets	8.33	3	Exceeds	11.75	6	Exceeds	11.17	6
UN-Habitat^	Approaches	7	6	Approaches	6.33	3	Meets	10	6
UNFPA*	Meets	9	11	Meets	8.87	15	Meets	9.62	13
UNODC*	Approaches	4	6	Approaches	4.36	11	Approaches	6.53	19
ITC	Did not use scorecard			Meets	10.4	5	Exceeds	12	4

³⁰ Entities that did not use the scorecard in 2014 or 2015 are not included in the analysis presented in the report.

UNCDF^	Meets	7.66	1	Meets	9	2	Exceeds	11	2
UNRWA^	Approaches	6.25	6	Exceeds	12	1	Approaches	6.8	5
DSS	Did not use scorecard			Meets	10	1	Missing	3	1
ESCAP^	Did not use scorecard			Meets	8.75	4	Meets	9.88	8
OIOS	Exceeds	10.85	6	Exceeds	10.83	6	Approaches	7.43	7
WHO	Meets	4.5	2	Meets	9.33	3	Meets	7.6	5
WIPO	Approaches	4.67	3	Meets	8.67	3	Meets	8	4
UNIDO^	Meets	7.78	18	Did not use scorecard			Meets	7.57	30
PBSO	Approaches	6	3	Did not use scorecard			Exceeds	11.67	3
UNITAR	Did not use scorecard			Did not use scorecard			Approaches	7	3
UNAIDS	Did not use scorecard			Did not use scorecard			Meets	8	3
GEF	Missing	2.6	5	Approaches	4	6	Approaches	4.43	7

^Participated in PLE in 2016

*External Review in 2016