



UNICEF Global Evaluation Report Oversight System

Quality Review of 2010 Evaluation Reports

Final Report v 1.3

Prepared for //
UNICEF Evaluation Office

Date // November 2011

By// Joseph Barnes
Hatty Dinsmore
Sadie Watson

IOD PARC is the trading name of International
Organisation Development Ltd//

Omega Court
362 Cemetery Road
Sheffield
S11 8FT
United Kingdom

Acknowledgements

This is the second global meta-evaluation report to use the current format and methodology. It continues to be possible thanks to the on-going support, feedback and contributions of Marco Segone and Abigail Taylor in the UNICEF Evaluation Office. Valuable and detailed feedback was also received from a number of UNICEF regional offices, and this proved to be vital in refining the final product.

We would like to appreciate the inputs of the evaluation reviewers, especially the speed with which new reviewers were able to integrate themselves into the team. At the heart of this meta-evaluation, Hatty Dinsmore continued to play a critical role in managing all of the parallel review, data management, and analytical processes; as well as adding substantially to the review matrix. Significant thanks are also due to Judith Friedman for courageously stepping into this role.

Joseph Barnes and Sadie Watson, November 2011

Acronyms

CCC	Core Commitments to Children
CEE/CIS	Central and Eastern Europe/Commonwealth of Independent States (Regional Office)
CFS	Child Friendly Schools
CO	Country Office
EAPRO	East Asia and Pacific Regional Office
EQOS	(UNICEF) Evaluation Quality Oversight System
ESARO	East and Southern Africa Regional Office
GEROS	(UNICEF) Global Evaluation Report Oversight System
GEWE	Gender Equality and Women's Empowerment
HRBAP	Human Rights Based Approach to Programming
IEC	Information, Education and Communication (materials)
M&E	Monitoring and Evaluation
MENA	Middle East and North Africa (Regional Office)
MfDR	Managing for Development Results
MTSP	Medium Term Strategic Plan
N/A	Not Applicable
RBM	Results Based Management
RO	Regional Office
ROSA	Regional Office of South Asia
TACRO	The Americas and Caribbean Regional Office
TOR (ToR)	Terms of Reference
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
WASH	Water, Sanitation and Hygiene
WCARO	West and Central Africa Regional Office

Executive Summary

UNICEF holds a long-standing commitment to independent assessment of the quality of evaluation reports produced by its country and regional offices all over the world, as well as HQ divisions. This is the second report to use the current methodology; developed in response to a need to provide decision makers in UNICEF with information about evaluation reports that better supports using and improving the knowledge generated by the evaluation function. It is important to note, however, that this approach is limited in its scope for achieving this aim because it is **only able to assess the quality of reports, and not of the evaluations behind them.**

This quality review process covered all 2010 evaluation reports submitted to the UNICEF Global Evaluation Database by the cut-off date of end of March 2011. The standards against which evaluation reports are assessed are set by the UNICEF deployment of the United Nations Evaluation Group (UNEG) global evaluation report standards.

Specific objectives of the review are to:

- Review and rate (with justifications) the quality of the main elements of evaluation reports, including structure, context, purpose, methodology, findings, conclusions, recommendation and lessons learned;
- To provide constructive feedback for evaluation commissioners to improve future evaluations;
- To provide a global analysis of key trends, strengths, weaknesses, and lessons of UNICEF evaluation reports; and
- To provide actionable conclusions and recommendations to improve the quality oversight system and systemic quality of the evaluation function;
- To provide specific information on the extent to which equity issues are taken into consideration.

Methodology

Evaluations were filtered out from other types of report. This process resulted in 89 full reviews. An evaluation expert who is familiar with the UNICEF evaluation

function and who had completed a dedicated induction process undertook each review. Three levels of quality assurance were applied: basic completeness and grammar checks; sampled peer-reviewing; and a right-to-challenge option exercised by the UNICEF Evaluation Office.

The full review tool is presented in the Annexes. This was originally co-designed by UNICEF and IOD PARC in 2010 based upon the UNEG/UNICEF standards. Changes made for this review include having separate questions on human rights, gender and equity; collecting more data on the role of TORs in evaluation quality; and reorganising the 58 guiding questions around 22 sub-sections. These sub-sections are used to generate a performance ‘dashboard’ presented in the annexes.

Each question, each section and the overall report are given a rating of either: ‘very confident’, ‘confident’, ‘almost confident’, or ‘not confident’ (where relevant, a N/A option is also provided). In addition to ratings, commentary is provided against each section and sub-section, suggestions for future improvement provided for each section, and executive feedback provided for each section and the overall report.

The review process generated an extensive dataset to inform the trend analysis, consisting of 1,068 individual pieces of evaluation typology data, 5,963 individual ratings, and 3,738 sections of qualitative text (approximately 120,000 words). In order to distil the key findings from this data, a multi-stage process was adopted consistent with the previous meta-evaluation.

This review added two additional layers of analysis that were possible because of time saved in the data aggregation process (after lessons learned in the previous meta-evaluation). The first additional analysis was a manual affinity diagramming of all of the overall comments made by reviewers. This was able to reveal meta themes, or clusters, that were not word-specific. The second additional layer of analysis was the quantitative cross-referencing of ratings data, from example between ‘equity’ and ‘MTSP correspondence’ in order to visually identify possible explanations for levels of report quality.

The limitations of time on the level of data analysis were mitigated as far as possible through triangulation of quantitative and qualitative patterns in the data. As noted previously, however, this does not enable us to overcome the limitations of only assessing the evaluation report and not the evaluation itself. Furthermore, the approach adopted is limited to being able to identify only the ‘headline’ findings, with

the possibility that more nuanced or infrequently occurring issues exist for individual readers to find within the reviews themselves.

Findings

This report draws on 89 reports from the following regions:

TACRO	9	MENARO	16
CEECIS	8	ROSA	4
EAPRO	10	WCARO	12
ESARO	21	Corporate	9

Overall, the review found a slight improvement in performance year-on-year, with 40% of reports being rated overall as satisfactory according to UNICEF Evaluation Standards (36% in 2010). Of the remaining reports, 30% could be brought up to a satisfactory standard with a little more work. Of some concern, however, was the 30% of reports found to have substantive weaknesses: 14% were in the previous year. At the top end of the scale, three reports were rated as outstanding¹ – although eight reports had at least one individual section rated as outstanding.

Our qualitative analysis suggests that there is a general trend of reports rated unsatisfactory as being highly descriptive in nature; as compared to better performing reports that used a much more analytical approach. This belies the fact that the differentiating factors between stronger and weaker reports are still fundamental evaluation issues: criteria, methodology, conclusions and so on.

There are flourishes of reports that deal innovatively with things such as stakeholder engagement or future policy. In general, however, well-structured reports do better simply because they develop a clear line of logic that results in evaluative conclusions based upon evidence, and there are a significant number of evaluation reports that still do not achieve this.

¹ Forward looking evaluation of the Community Schools Project UNICEF, Egypt; Cash Transfer Programme for Orphans and Vulnerable Children Operational and Impact Evaluation, Kenya; Evaluation of the Go to School Initiative in Southern Sudan, Sudan.

Thus the overriding theme coming out in this meta-evaluation is the influence of clear and sufficient evidence in determining the quality rating of evaluation reports. This requirement for clear presentation, robust analysis, and transparent interpretation of information cuts across all of the assessed sections: from the selection of evaluation criteria to the development of lessons learned.

Reports were found to be stronger in terms of overall coherence (45% confidence) than in terms of individual sections (35% confidence). This would suggest that in overall terms, individual sections are causing more problems than poor structuring.

Far more 2010 reports included Terms of Reference: 64% compared to 34% of 2009 reports. This is to be commended. Of those reports that did have a TOR attached, around 36% appeared to have a broadly positive effect on the final report quality.

The most consistently satisfactory reports were found to be global and regional evaluations. Corporate global evaluations, in particular, were found to be almost entirely satisfactory in the critical methodology and conclusions sections that let down other reports.

The consistently strongest performing reports were those managed jointly by UNICEF and another UN agency: although this was a fairly small number of reports (five), 80% were found to be satisfactory. UNICEF-managed reports were often more challenging impact evaluations, whereas Joint-UN reports tended to focus at the outcome level. Country-led evaluations were found to be the weakest, with only 25% rated as reaching UNICEF standards.

Policy evaluations were found to be very few in number (three) and generally of unsatisfactory quality according to UNICEF standards. Real time humanitarian evaluations – also a small sample (seven) – were found to be consistently strong.

Output evaluations were by far the least satisfactory (less than 20% rated satisfactory), outcome evaluations were stronger but suffered from weak methodology sections (32% satisfactory), and impact evaluations were strongest (66% rated satisfactory).

Multi-sector evaluations were found to be the strongest overall, with 56% rated satisfactory. Young Child Survival (MTSP1) and Basic Education (MTSP2) evaluation reports were most consistently rated as satisfactory among single-sector reports.

Independent Internal evaluations were rated higher (55%) than Independent External evaluations (41%). There were no self-evaluations.

Quantitative analysis of section ratings suggests that the greatest correlation (80%) between a section rating and the overall rating is the findings and conclusion section. Outstanding reports tended to be differentiated by the methodology section. Developing a clear chain of logic from data to evaluative conclusions is the major barrier to many reports reaching UNICEF standards

The object of the evaluation was one of the highest rating sections, with 46% of reports being placed as confident or very confident to act. Purpose, objectives and scope was the second strongly rated section, with 47% of reports reaching UNICEF standards. Methodology was generally found to be an area of some concern across all evaluation reports, with 65% of reports rated as below standard.

At 63% below standards, the findings and conclusions section showed some considerable concern, and was the weakest section in reports rated as unsatisfactory overall. Findings and lessons learned was the weakest performer, with only 34% of evaluations being found to satisfactorily meet UNICEF standards: due mostly to misinterpreted lessons learned. Finally, 42% of reports rated as satisfactory in relation to the report structure, logic and clarity.

As part of this meta-evaluation, a special analysis of the performance of evaluation reports in relation to equity was requested. Overall, only 9% of reports included equity issues to a satisfactory level, according to UNICEF standards: one report was rated outstanding and seven were rated as confident. A very strong correlation was found in the level of coverage of equity issues by reports and their overall quality rating. Real Time Evaluations handled equity better than any other purpose.

Jointly managed evaluations with UN agencies and country-led evaluations were both small in number (five each), but produced proportionately higher rates of satisfactory reports (both 20%) and lower rates of not confident ratings (20% and 40% respectively).

Quantitative analysis would appear to suggest that performance in equity is largely dependent on individual combinations of evaluators and evaluation managers, rather than an institutional capacity located in some part of the organisation: a major challenge to systemic strengthening.

The differentiating factor for reports that satisfactorily addressed equity was their consistency in mainstreaming equity considerations throughout the evaluation: from design, framework and methodology through to data collection, analysis and conclusions. Most reports rated ‘almost confident’ offered consideration of equity as an issue, some few actually collected disaggregated data, and very few managed to generate convincing analysis of this.

Whilst human rights was better covered than equity – with 19% of reports being rated satisfactory – no evaluation was found to be outstanding in this regard. Similarly for gender equality and women’s empowerment issues: 20% of reports were rated as satisfactory, but none were rated outstanding.

The level of confidence in ethics sections has substantially declined since 2009, from 18% to only 10% of reports reaching the (liberally interpreted) UNICEF standards. The degree to which ethics is discussed now appears to be on life-support, and warrants renewed highlighting of this concern.

Regional analysis reveals a mixed performance compared to the previous year. There was slight overall improvement, but a concerning increase in the number and proportion of reports rated as *not confident*. Tracking the regional distribution of performance over the two years 2009-2010 shows that, with the exception of Corporate evaluations, the relative ordering of regions as most-to-least satisfactory reports performed an almost precise reversal. Nevertheless, all except two regions had at least one report where a section was rated outstanding.

There does appear to be a pattern emerging of French language reports rating less well than all of the regional averages from which they came. To be able to assert such a conjecture conclusively, however, is beyond the scope of this methodology.

Conclusions

Conclusions were developed by analysing the findings for trends in underlying factors that contributed to the performance of evaluation reports. This analysis was grounded in the concept of ‘confidence’.

Inconsistency of ratings across time, regions and sectors suggests low institutionalisation of quality evaluation capabilities

It would appear that – with the exception of corporate evaluations – there are not really ‘stores’ of capabilities concentrated anywhere in the system. Great evaluation reports seem to be just as likely to come out from one point in the evaluation system as any other: as do very poor quality reports.

UNICEF appears not to be using its comparative advantage in equity, gender and human rights

UNICEF has been at the forefront of international efforts to recognise human rights, gender equality, women and children’s empowerment, and socio-economic equity within the evaluation function. This is not being translated into strong performance in these areas.

The performance of impact evaluations and joint UN evaluations could suggest higher resourcing of evaluations is leading to better reports – but might not

On the assumption that impact evaluations and joint evaluations tend to involve higher levels of management time and resourcing, the findings could imply that higher levels of investment are generally leading to better quality reports. Alternatively, jointly-managed evaluations are simply delivering well on easier outcome-level questions, whereas UNICEF is managing inherently more complex and challenging evaluations that others do not want to get involved in.

Evidence, logic and analysis really counts in delivering quality

Outstanding reports are differentiated by the methodology, whereas other reports are most influenced by whether they get the findings-conclusions nexus right. Focusing on the broad detachment between evaluation design and objects' theories of change could help to move from reports full of description to those containing more analysis and logic. Most reports that are sub-standard appear confused because they leave important issues about the evaluation design or implementation unexplained.

The purpose of Lessons Learned is not understood

As a knowledge-centred organisation, it has to be of concern to UNICEF that only 15% of reports adequately demonstrate an understanding of what the purpose and value of lessons learned is. Lesson learned are mostly misinterpreted as object-specific management changes.

Ethics is an issue on life-support

Whilst no conclusions can be drawn regarding how ethical evaluations actually are, the vast majority of evaluation reports simply fail to even mention ethics issues.

Lessons Learned

Lessons learned have been generated through the analysis of the core-evaluation team and revised based on comments from UNICEF.

Evaluation reports need evaluation skills, but they also need proactive inputs from specialists in rights, gender, ethics and equity.

Evaluations undertaken by technical experts in a research style result in weaker quality of reports than those undertaken by evaluators. It is also the case, however, that most evaluators appear to be unsuccessful in building an analytical framework for their reports around human rights and gender commitments. This suggests a strong need for rights experts to be closely involved with evaluation teams – particularly at the design stage.

Conclusions from a single year's meta-evaluation cannot give the whole picture.

Taken in isolation, there are a number of trends apparent in the 2009 and 2010 reports. Several of these, however, do not hold across the two years' reviews. The lessons from this experience are to exercise caution in interpreting trends in a single

meta-evaluation, and to ensure that meta-evaluation methodologies continue to be consistent across years.

Recommendations for the Evaluation Office

The following recommendations have been generated through the analysis of the core-evaluation team.

Explore the option of a developing a specific enhancement strategy for policy evidence

With an organisational commitment to becoming a global knowledge provider, policy evaluations and lessons learned can be seen as critical business areas for the evaluation function, and may warrant a specific strategy to accelerate their enhancement.

Consider putting communications at the centre of a campaign approach to ethics and equity

There would appear to be an opportunity for UNICEF to apply some of its communications capability internally to influence the behaviour of evaluation managers and suppliers. The two issues that stand out for such treatment are equity and ethics.

Consider investigating the real effects of the language gap

Two consecutive meta-evaluations have found reason to infer that non-English language evaluation reports may be hampered by disparities in volume, timeliness and conceptual nativeness of evaluation guidance. It would seem to be worth decisively exploring the real nature of the effect of language, if any, on evaluation quality.

Recommendations for Regions and Countries

Explore strategies for developing a cohort of affordable evaluators specifically trained in the needs of UNICEF evaluation reports

As the evaluation manager, UNICEF is ultimately able to exercise substantive influence over the quality of the final report and of the evaluation itself. A complementary strategy to developing internal capacity would also be to develop long term relationships with affordable local evaluators that would allow them to invest in the specific capabilities required to meet UNICEF's needs. This could include a

training course that provided detailed insights into UNICEF standards and how they are expected to be applied.

Consider developing a pool of experts in human rights, gender, equity and ethics who can advise evaluation teams

UNICEF may wish to consider developing a pool of expertise that is available to advise evaluation teams during the inception phase of evaluations. This would help to strengthen the integration of rights, gender, equity and ethics into contextualised evaluation frameworks.

Terms of Reference show progress, but would benefit from clear tables of contents and direct hyperlinks to evaluation standards

Including a draft table of contents in the terms of reference – with headings such as ‘evaluation framework’, ‘limitations of the methodology’, ‘ethical consideration’, and ‘lessons learned’ – not only helps to guide the report writers, but can also act as an easy administrative check in signing off the report.

Maximise the value of the inception period by introducing a pro-forma report where evaluators explain how they intend to implement UNICEF guidance, analyse evidence and develop recommendations

Reports rate well that have a clear and transparent methodology (not necessarily spectacular or innovative, just clear), a set of transparently processed evidence (i.e. a presentation of findings rather than data), and a clear set of evaluative conclusions based on this evidence. They also meet all of UNICEF’s guidance on specific issues (such as equity). By including an inception period, evaluators have the opportunity to familiarise themselves with the object and thus better propose how UNICEF standards will be realised in that context. This proposal could be presented in a standardised format that outlined a) how the evaluation team interpreted specific pieces of UNICEF guidance, and b) how these will be realised in each specific case.

Contents

Executive Summary	iii
Review Background and Purpose.....	1
<i>Scope of the review</i>	<i>1</i>
<i>Objectives of the review</i>	<i>2</i>
<i>Specific objectives.....</i>	<i>2</i>
Review Methodology	2
<i>UNICEF Evaluation Report Standards</i>	<i>3</i>
<i>Review tool, process and analysis.....</i>	<i>4</i>
Report typology	5
Assessment ratings	5
Justification, feedback and recommendations.....	6
Meta-evaluation Methodology.....	7
<i>Trend analysis</i>	<i>7</i>
<i>Limitations of the approach</i>	<i>8</i>
Findings.....	9
<i>Overall findings on quality.....</i>	<i>9</i>
Coherence of the reports	10
Geography.....	11
Management.....	11
Purpose.....	12
Result.....	13
MTSP correspondence.....	13
Level of independence	14
Stage/timing	14
<i>Section Findings.....</i>	<i>15</i>
Section A: Object of the evaluation	16
Section B: Evaluation purpose, objectives and scope	17
Section C: Evaluation methodology	18
Section D: Findings and conclusions	20
Section E: Recommendations and lessons learned.....	21
Section F: Report structure, logic and clarity.....	22
<i>Special Section: Findings on Equity.....</i>	<i>24</i>
Gender and human rights	29
Ethics	29
Participation	30
<i>Regional findings.....</i>	<i>31</i>
Language differences.....	31
Examples of interesting and innovative practices.....	32
Conclusions	36
<i>Lessons Learned</i>	<i>39</i>
Recommendations	41
<i>Recommendations for UNICEF Evaluation Office</i>	<i>41</i>
<i>Recommendations for Regional and Country Offices</i>	<i>42</i>
Annexes	45
<i>Annex 1: Terms of Reference</i>	<i>45</i>
<i>Annex 2: Review Team</i>	<i>47</i>

<i>Annex 3: List of Assessed Reports</i>	<i>49</i>
<i>Annex 4: Review Tool</i>	<i>53</i>
<i>Annex 5: Regional Ratings Graphs</i>	<i>59</i>
<i>Annex 6: Ratings Graphs by Typology</i>	<i>62</i>
<i>Annex 7: Performance Dashboard</i>	<i>65</i>
<i>Annex 8: Additional Equity Graphs</i>	<i>69</i>
<i>Annex 9: Classification of reviewed 2010 reports</i>	<i>72</i>

Figures and Tables

Figures

<i>Figure 1: distribution of overall ratings for 2010 evaluation reports</i>	<i>9</i>
<i>Figure 2: word cloud of reviewer's overall comments</i>	<i>10</i>
<i>Figure 3: the broad positive or negative effect that TORs were perceived to have on report quality (where they were included).</i>	<i>11</i>
<i>Figure 4: numbers of reports of different purpose by management type</i>	<i>12</i>
<i>Figure 5: percentage matching between section ratings and the overall rating of a report</i>	<i>15</i>
<i>Figure 6: Spread of results for Section A</i>	<i>17</i>
<i>Figure 7: Spread of results for Section B</i>	<i>18</i>
<i>Figure 8: Spread of results for Section C</i>	<i>19</i>
<i>Figure 9: Spread of results for Section D</i>	<i>20</i>
<i>Figure 10: Spread of results for Section E</i>	<i>22</i>
<i>Figure 11: Spread of results for Section F</i>	<i>23</i>
<i>Figure 12: the inclusion of equity issues in reports</i>	<i>25</i>
<i>Figure 13: the coverage of equity by different management arrangements</i>	<i>27</i>
<i>Figure 14: discussion of ethics in 2009 and 2010 reports</i>	<i>30</i>
<i>Figure 15: spread of report ratings by region</i>	<i>31</i>

Tables

<i>Table 1: proportion of different reports rated confident by language</i>	<i>32</i>
<i>Table 2: examples of good evaluation practice noted in the regions:</i>	<i>32</i>
<i>Table 3: examples of good programme practice identified in reports</i>	<i>34</i>

Review Background and Purpose

UNICEF holds a long-standing commitment to independent assessment of the quality of evaluation reports produced by its country and regional offices all over the world, as well as HQ divisions. Most notably, this has included large-scale meta-evaluations, including Victora et al (1995), Watson et al (2004), Noij et al (2009) and Barnes et al (2010).

The quality review that is the subject of this report sits within this tradition. It is the second report to use the current methodology; developed in response to a need to provide decision makers in UNICEF with information about evaluation reports that better supports using and improving the knowledge generated by the evaluation function.

This qualitative approach seeks to actively inform the enhancement of evaluation report quality. Furthermore, it seeks to go beyond raising awareness of quality issues, and demonstrate the implications of trends in evaluation quality on the usability of knowledge in the pursuit of delivering results for children and women. It is important to note, however, that this approach is limited in its scope for achieving this aim because it is **only able to assess the quality of reports, and not of the evaluations behind them.**

Scope of the review

This quality review process covered all 2010 evaluation reports submitted to the UNICEF Global Evaluation Database by the cut-off date of end of March 2011. Reviews, research and other types of reports were excluded, as were evaluation reports undertaken by offices or divisions that were not submitted to the UNICEF Global Evaluation Database.

The quality review tool assesses the evaluation report as a *standalone* document. No additional investigation is made into the use, conduct or context of an evaluation. The standards against which evaluation reports are assessed are set by the UNICEF deployment of the United Nations Evaluation Group (UNEG) global evaluation report standards.

Objectives of the review

The overall objective is to assess and rate the quality of evaluation reports commissioned by UNICEF in 2010 using the UNEG/UNICEF Evaluation Report Standards.

Specific objectives

- Review and rate (with justifications) the quality of the main elements of evaluation reports, including structure, context, purpose, methodology, findings, conclusions, recommendation and lessons learned;
- To provide constructive feedback for evaluation commissioners to improve future evaluations;
- To provide a global analysis of key trends, strengths, weaknesses, and lessons of UNICEF evaluation reports; and
- To provide actionable conclusions and recommendations to improve the quality oversight system and systemic quality of the evaluation function;
- To provide specific information on the extent to which equity issues are taken into consideration.

Review Methodology

The meta-evaluation considered all of the screened reports that had been submitted as evaluations to the UNICEF Global Evaluation Database for 2010. These reports were subjected to an initial filtering process by an evaluation expert to remove from the sample frame those reports that could easily be considered as being a form of assessment other than evaluation according to UNICEF standards.

“An Evaluation is judgment made of the relevance, appropriateness, effectiveness, efficiency, impact and sustainability of development efforts, based on agreed criteria and benchmarks among key partners and stakeholders. It involves a rigorous, systematic and objective process in the design, analysis and interpretation of information to answer specific questions. It provides assessments of what works and why, highlights intended and unintended results, and provides strategic lessons to guide decision-makers and inform stakeholders.”
(UNICEF, 2011)

Reports were further filtered according to this definition by each reviewer prior to review. This process resulted in 89 full reviews. Each review was undertaken by an evaluation expert familiar with the UNICEF evaluation function. All reviewers had either participated in the previous global meta-evaluations, or went through a peer-reviewed induction process to enable a common understanding of the standards to be reached based upon practical experience.

Following the completion of each review format, three further levels of assurance were applied. The first was a basic ‘completeness’ check to ensure that all relevant information had been provided. Secondly, a peer review by a member of the core evaluation team was undertaken on selected reports for each reviewer, with any disparities between results flagged to the original reviewer for clarification and/or amendment. Finally, the UNICEF Evaluation Office were able to exercise the right to challenge any review on technical grounds, with comments provided to the original reviewer for consideration.

UNICEF Evaluation Report Standards

In line with its international commitments, UNICEF has adopted a contextualised version of the United Nations Evaluation Group evaluation report standards. These identify 42 standards arranged under 8 main sections: report structure, object of the evaluation, evaluation purpose objectives and scope, evaluation methodology, findings, conclusions and lessons learned, recommendations, and gender and human rights.

These standards inform the UNICEF Global Evaluation Report Oversight System. One of the four main purposes of the Global Evaluation Report Oversight System (GEROS) is to assess the quality of evaluation reports: the component under which this meta-evaluation sits. GEROS establishes the standard against which this meta-evaluation assesses evaluation reports to be satisfactory or not.

“An evaluation report is assessed as satisfactory when it is a credible report that addresses the evaluation purpose and objectives based on evidence, and therefore can be used with confidence.” (UNICEF, 2010 p3)

In addition, the GEROS provides the basis for the development of a review tool by establishing five core ‘elements’ of an evaluation report that are required to respond to the eight adapted UNEG standards:

- ☐ Well structured, logical and clear
- ☐ Clear and full description of the 'object' of the evaluation
- ☐ The evaluation's purpose, objectives and scope are fully explained
- ☐ Appropriate and sound methodology
- ☐ Findings, conclusions and recommendations are based on evidence and sound analysis

Review tool, process and analysis

The full review tool is presented in the Annexes. This was originally co-designed by UNICEF and IOD PARC in 2010 based upon the UNEG/UNICEF standards. Changes made for this review include having separate questions on human rights, gender and equity; collecting more data on the role of TORs in evaluation quality; and reorganising the 58 guiding questions around 22 sub-sections. These sub-sections are used to generate a performance 'dashboard' presented in the annexes.

The review tool primarily adopts a qualitative approach to rating evaluation reports against the overall standard of confidence. It pursues a systematic process of aggregating qualitative ratings of the guiding questions about different aspects of an evaluation report into six sections, and then into a final overall assessment. The six sections of analysis are based on the five core elements identified in the Evaluation Quality Assurance System, with the further separation of recommendations and lessons learned as a separate lens of enquiry. The six sections of the review tool are:

- ☐ Object of the evaluation
- ☐ Purpose, objectives and scope
- ☐ Evaluation methodology, gender, human rights and equity
- ☐ Findings and conclusions
- ☐ Recommendations and lessons learned
- ☐ Report is well structured, logical and clear
- ☐ (Plus additional information and an overall reaction)

This qualitative approach is designed to enable reviewers to provide useful analysis across the range of evaluation contexts encountered; and constructive feedback to improve future evaluation reports. Each question, each section and the overall report

are given a rating of either: ‘very confident’, ‘confident’, ‘almost confident’, or ‘not confident’ (where relevant, a N/A option is also provided).

In addition to ratings, commentary is provided against each section and sub-section, suggestions for future improvement provided for each section, and executive feedback provided for each section and the overall report. The complete review process generates three types of data: a report typology, a series of ratings, and a structured set of discussion text.

Report typology

Evaluation reports are initially classified according to commissioning agent and the UNICEF evaluation typology. This allows analysis according to various evaluation report characteristics. The seven typology criteria are: geographical coverage, management, purpose, result level, MTSP correspondence, level of independence, and timing.

Assessment ratings

Assessment ratings were given against 58 guiding questions, six sections, two meta-questions, and an overall view of the report. Each ‘aspect’ was given one of four qualitative ratings: two ratings accorded to the reviewer finding that a UNICEF manager could confidently act on that aspect of the evaluation report, and two ratings accorded to the finding that a UNICEF manager could not be confident in that aspect of the evaluation report.

Each rating was informed by three factors: a prompting question in the review tool; a ‘confidence-to-act’ test, and any ratings in the level of analysis below. In other words, the overall rating was informed by the six section ratings, and each section rating was informed by the question ratings clustered under that section. Within this ‘aggregating’ system, each individual rating had the following properties:

1. **Ratings are absolute measures.** *Each report rating is made against the UNICEF Evaluation Report Standards. It is therefore possible for all reviewed reports to achieve an ‘outstanding’ rating, or all reports to be classified with a ‘not confident’ rating. There is no relative assessment within a cohort of reports.*
2. **Ratings are not compoundable.** *Ratings are not ‘quantitated’ representations of qualitative assessments and cannot be used for calculating ‘mean’ ratings. Thus, a particular report may have a mix of*

‘outstanding and ‘confident’ section ratings, but these do not automatically add up to give a particular overall rating. Rather, the different levels of ratings (questions, sections, and overall) allows the reviewer to build up a holistic overall view on the quality of an evaluation report. Similarly, overall report ratings are not compoundable: thus, a particular region cannot be given an ‘average’ score.

- 3. Ratings are not distributed on an even scale.** *There are four possible ratings available to each reviewer. The main distinction between these is between ‘confident-to-act’ and ‘not confident’. An additional two ratings – ‘outstanding’ (or ‘very confident’) and ‘almost confident’ – allow a reviewer to provide a more nuanced response to highlight a narrow selection of reports (or elements of reports) that can be flagged as either best practice examples, or requiring just minor amendments in order to reach a satisfactory standard.*

Justification, feedback and recommendations

The second main dataset to be generated is qualitative text linked to each rating. The primary purpose of this text is to enable evaluation managers to understand a) why their reports received particular ratings, and b) how to improve future reports. Executive summaries for each section and each report also provide rapid feedback to UNICEF senior managers that highlights the strengths, weaknesses and lessons for each report. Qualitative analysis of this rich dataset is used to inform the trend analysis.

Meta-evaluation Methodology

Trend analysis

The review process generated an extensive dataset to inform the trend analysis, consisting of 1,068 individual pieces of evaluation typology data, 5,963 individual ratings, and 3,738 sections of qualitative text (approximately 120,000 words). In order to distil the key findings from this data, the following multi-stage process was adopted.

1. All reviewers submitted analysis identifying the key themes (strengths, weaknesses and issues) that they had encountered in the evaluation reports that they reviewed, and in the review tool itself;
2. The core evaluation team undertook a manual affinity-mapping process for overall comments and synthesised reviewer feedback into the core issues identified under each section of the review tool;
3. All reviews were completed within an Excel template: these were then aggregated into a single database;
4. Rating data was disaggregated using evaluation typology data to produce a series of distribution bar graphs that visualised the spread of ratings for different aspects of the evaluation reports (examples of these can be found in the findings section);
5. Qualitative text comments were disaggregated according to overall report and section ratings and combined into blocks of text for analysis;
6. Blocks of text were subjected to inductive coding by using Word 'find and replace' tool and the Wordle common English words filter (such as "the", "and" and "a").
7. The inductive coding generated a list of most-frequently occurring words and issues disaggregated according to the ratings given. This was compared to the synthesis of reviewer comments in order to identify the prevalence of the main trends.
8. Additional quantitative analysis was conducted on specific questions, including cross-referencing multiple variables, in order to triangulate trends.
9. A selection of reports with representative or interesting ratings (or those flagged by reviewers) were then revisited in or to further triangulate the findings of the coding process, or to add substantiating details to these findings.

Changes from the previous year

This review added two additional layers of analysis that were possible because of time saved in the data aggregation process (after lessons learned in the previous meta-evaluation). The first additional analysis was a manual affinity diagramming of all of the overall comments made by reviewers. This was able to reveal meta themes, or clusters, that were not word-specific. The second additional layer of analysis was the quantitative cross-referencing of ratings data, from example between ‘equity’ and ‘MTSP correspondence’ in order to visually identify possible explanations for levels of report quality.

In addition to these analysis changes, the review questions were reordered to make them more logical, and separate questions were provided on human rights, gender and equity. Other than this, the review format and analysis was kept as consistent as possible in order to aid in identifying trends across time.

Limitations of the approach

As with all evaluations, the approach taken by this meta-evaluation is subject to real-world constraints, including resources and time. This requires a balance between what is ideal and what is possible: and creates a number of limitations that must be taken into account. Where relevant, these limitations have been mitigated through triangulation of both the review and trend analysis processes in order to constrain opportunities for inconsistency or information-loss. For instance, the review process itself is subject to the limitation of only having access to the written evaluation report (and these did not always include the Terms of Reference), with no reliable data on the evaluation cost, team or usefulness. As a consequence, findings and conclusions drawn can only be applied to an evaluation report, and not to the evaluation itself.

Qualitative analysis (as with all analysis) requires for judgements to be made in identifying the important indicators and trends contained within the dataset generated by the review process. An ideal approach may have been to separately consider each of the thousands lines of text written in order to cluster points around the main trends: unfeasible in this instance. Thus, the approach adopted is limited to being able to identify only the ‘headline’ findings, with the possibility that more nuanced or infrequently occurring issues exist for individual readers to find within the reviews themselves.

Findings

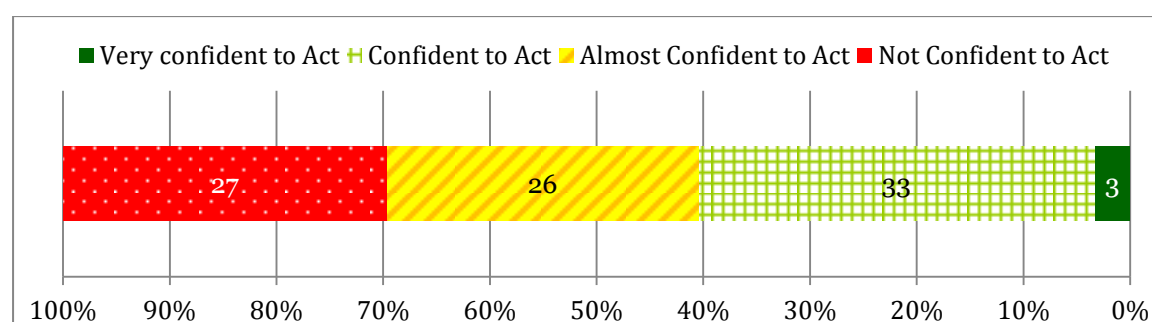
This report draws on 89 reports from the following regions:

TACRO	9	MENARO	16
CEECIS	8	ROSA	4
EAPRO	10	WCARO	12
ESARO	21	Corporate	9

Overall findings on quality

Overall, the review found a slight improvement in performance year-on-year, with 40% of reports being rated overall as satisfactory according to UNICEF Evaluation Standards (36% in 2010). Of the remaining reports, 30% could be brought up to a satisfactory standard with a little more work. Of some concern, however, was the 30% of reports found to have substantive weaknesses: 14% were in the previous year. At the top end of the scale, three reports were rated as outstanding² – although eight reports had at least one individual section rated as outstanding.

Figure 1: distribution of overall ratings for 2010 evaluation reports



Our qualitative analysis suggests that there is a general trend of reports rated unsatisfactory as being highly descriptive in nature; as compared to better performing reports that used a much more analytical approach. This belies the fact

² Forward looking assessment of the Community Schools Project UNICEF, Egypt; Cash Transfer Programme for Orphans and Vulnerable Children Operational and Impact Evaluation, Kenya; Evaluation of the Go to School Initiative in Southern Sudan, Sudan.

There are flourishes of reports that deal innovatively with things such as stakeholder engagement or future policy. In general, however, well-structured reports do better simply because they develop a clear line of logic that results in evaluative conclusions based upon evidence, and there are a significant number of evaluation reports that still do not achieve this.

Figure 2: word cloud of reviewer’s overall comments



Thus the overriding theme coming out in this meta-evaluation is the influence of clear and sufficient evidence in determining the quality rating of evaluation reports. This requirement for clear presentation, robust analysis, and transparent interpretation of information cuts across all of the assessed sections: from the selection of evaluation criteria to the development of lessons learned.

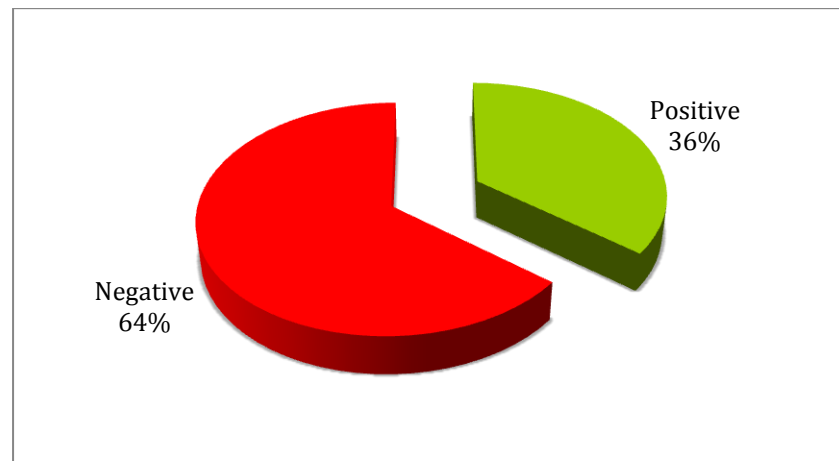
Coherence of the reports

As with the previous year, reports were separately rated in terms of both individual sections and overall coherence. As was the case with 2009 reports, 2010 reports were found to be stronger in terms of overall coherence (45% confidence) than in terms of individual sections (35% confidence). This would suggest that in overall terms, the picture remains the same of individual sections causing more problems than poor structuring in the higher-performing evaluation reports. Unlike the previous year, this seems to also be the case for poorly performing reports. There is more discussion on which sections appeared to be most influential under Section Findings, below.

Far more 2010 reports included Terms of Reference: 64% compared to 34% of 2009 reports. This is to be commended, as it enabled reviewers to gain a much better

impression on how TORs appeared to be influencing the quality of evaluation reports. Of those reports that did have a TOR attached, around 36% appeared to have a broadly positive effect on the final report quality; whereas, in 64% of cases, issues in the final report could be traced back to problems with the Terms of Reference.

Figure 3: the broad positive or negative effect that TORs were perceived to have on report quality (where they were included).



Geography

The majority of evaluation reports reviewed in this round – 62% – were at the sub-national level. These sub-national reports were also found to be the least satisfactory, with only just over 30% being rated as confident to act. National level reports, the second largest grouping, were found to be stronger: with 46% being rated as confident to act. However, these were generally prevented from being rated higher by weak methodology sections (fewer than 30% of national evaluations were rated as having satisfactory methodologies).

Consisting of a much smaller sample, the most consistently satisfactory reports were found to be global and regional evaluations. Corporate global evaluations, in particular, were found to be almost entirely satisfactory in the critical methodology and conclusions sections that let down other reports. There were no multi-country evaluations in this round.

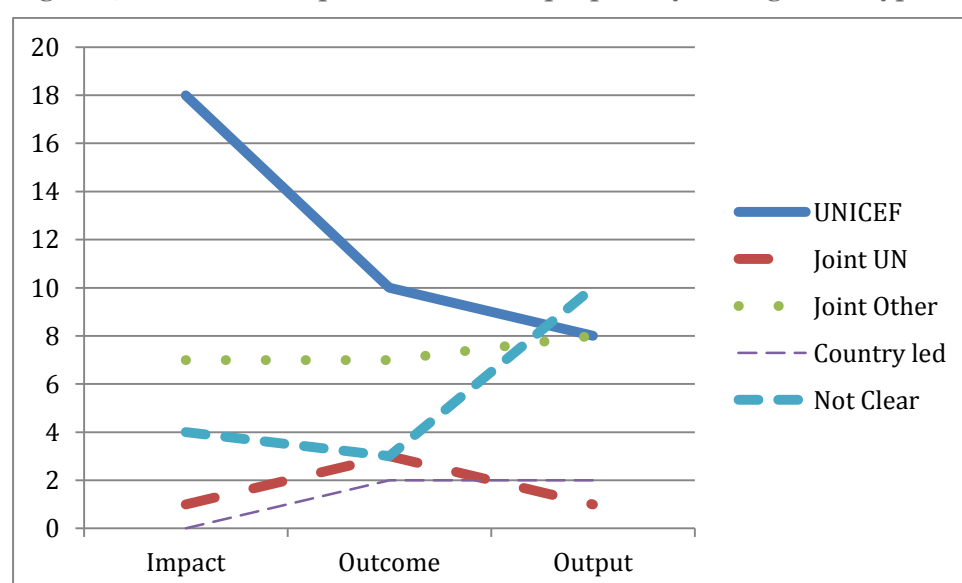
Management

For 15 of the 89 reports reviewed, it was not possible to tell what the management arrangements for the evaluation had been. All of these reports were also found to be unsatisfactory overall: which could suggest that poor ‘housekeeping’ in a final report is an early indicator that there are likely to be more fundamental problems within it.

The consistently strongest performing reports were those managed jointly by UNICEF and another UN agency: although this was a fairly small number of reports (five), 80% were found to be satisfactory. Country-led evaluations were found to be the weakest, with only 25% rated as reaching UNICEF standards. These reports particularly tended to have unsatisfactory methodology sections.

Of the remaining reports, UNICEF-managed evaluations were rated slightly more satisfactory than those managed jointly with other development partners; with both types tending to be weaker towards the final sections. It should be noted, however, that UNICEF-managed evaluation reports tended to be rated as having much better analytical consistency in findings and conclusions. In a cross-referencing of purpose and management it was found that UNICEF-managed reports were often more challenging impact evaluations, whereas Joint-UN reports tended to focus at the outcome level. This could be one explanatory factor for differing performance.

Figure 4: numbers of reports of different purpose by management type



Purpose

As with the previous meta-evaluation, policy evaluations were found to be very few in number (three) and generally of unsatisfactory quality according to UNICEF standards. Thus, this remains an area of concern in relation to our understanding of UNICEF's upstream strategy. Conversely, real time humanitarian evaluations – also a small sample (seven) – were found to be consistently strong; with over 70% rated satisfactory and suggesting that UNICEF maintains this core value-adding capacity in humanitarian response.

From the larger groupings, project evaluations were found to be the weakest, with less than 30% being found to be confident to act. It would seem to be reasonable to conjecture that this is consistent with the findings (below) that output evaluation reports are also weak, and that this may be a consequence of lower investment (in both evaluators and management) compared to other reports.

Despite 60% programme evaluations being rated as unsatisfactory overall, 60% of these could have been brought up to UNICEF standards fairly easily. By comparison, evaluations of 'at-scale' programmes were stronger overall, but had a higher proportion rated as fully not-confident to act (almost 40%). From the larger sample-size groups, pilot evaluations were the strongest, perhaps reflecting their emphasis on evidence generation for scaling up models.

Result

There was fairly even distribution of evaluation reports across the levels of result – output, outcome, impact – which they sought to evaluate. Furthermore, there was a very clear trend in ratings: output evaluations were by far the least satisfactory (less than 20% rated satisfactory), outcome evaluations were stronger but suffered from weak methodology sections (32% satisfactory), and impact evaluations were strongest (66% rated satisfactory). From this, it seems to be not unreasonable to surmise that there is a strong causal relationship between the level of investment in an evaluation and the quality (and usefulness) of the report that is generated.

MTSP correspondence

Multi-sector evaluations – evenly spread across output, outcome and impact – were found to be the strongest overall, with 56% rated satisfactory and the best methodology sections. From the MTSP focus areas, Young Child Survival (MTSP1) and Basic Education (MTSP2) evaluation reports were most consistently rated as satisfactory, with around half being rated as confident to act (41% and 50% respectively). Both, however, also had concerning levels (over 37%) of reports being rated as not confident; and also rated poorly in relation to methodology.

Child Protection (MTSP4) evaluations were also found to have issues relating to methodologies and rated weaker overall than education and young child reports, but had less than 30% rated as not confident – suggesting that it could improve more easily. Only one report – an impact evaluation – was classified as being Policy and Advocacy (MTSP5), and this was rated as not confident: reflecting the concern identified above. Cross-cutting evaluations were mostly at output level, and

consequently largely rated unsatisfactory; again methodology sections were found to be critical.

HIV/AIDS evaluation reports (MTSP3) were also found to be weak, particularly in relation to the findings and conclusions section: although it should also be noted that one HIV/AIDS report was rated as outstanding. In the previous meta-evaluation it was conjectured that this may be caused by a prevalence of output level HIV/AIDS evaluations; however in cross-checking this year's reports it was found that there were the same number of HIV/AIDS impact evaluations as output evaluations (four each).

Level of independence

Within the reports reviewed, there were no self-evaluations: highlighting the value placed on independence by evaluation commissioners. Independent Internal evaluations were rated higher (55%) than Independent External evaluations (41%). This difference is more pronounced in earlier sections, both dropping to around 40% satisfactory for lessons learned. Of some concern, the independence of nine evaluations could not be determined from their reports.

Stage/timing

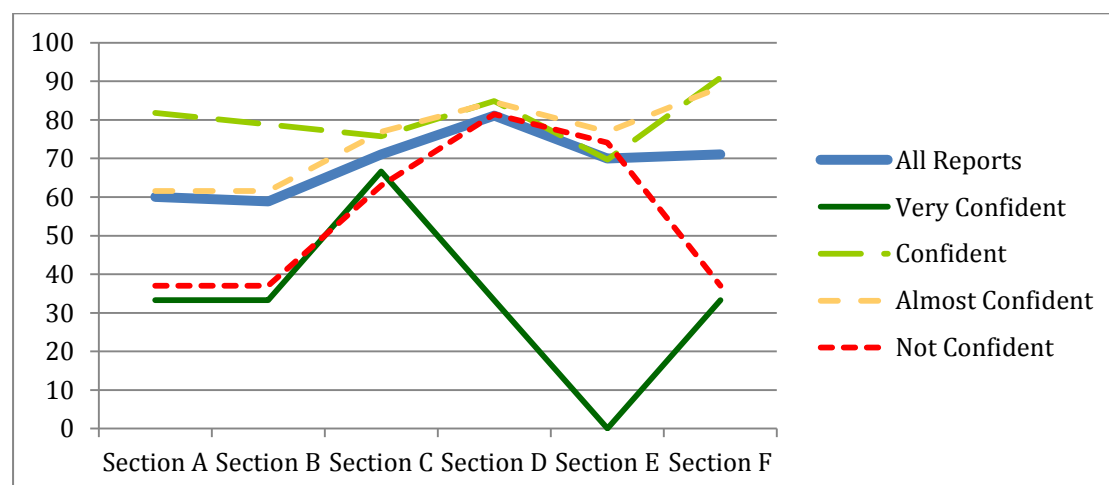
Our analysis found no significant difference between Summative and Formative, with similar numbers in both across all sections.

Section Findings

In line with the language in UNICEF standards, reports are referred to overall in terms of being *satisfactory* or not; and individual sections according to the test of *confidence*. Quantitative analysis of section ratings suggests that the greatest correlation (80%) between a section rating and the overall rating is the findings and conclusion section. For satisfactory and almost satisfactory reports, the structure section (including executive summary) was highly correlated with overall quality, but this was much less influential on the extremities of outstanding and not confident reports. Outstanding reports tended to be differentiated by the methodology section.

Considering the spread of ratings separately for reports rated satisfactory and reports rated as unsatisfactory reveals that the methodology and lessons learned sections were the areas most in need of improvement among overall satisfactory evaluations. Whilst these were also areas of concern among reports rated as unsatisfactory, the findings and conclusions section is what appears to be most problematic for these evaluations. This supports the qualitative finding, above, that developing a clear chain of logic from data to evaluative conclusions is a major barrier to many reports reaching UNICEF standards.

Figure 5: percentage matching between section ratings and the overall rating of a report



Section A: Object of the evaluation

“The report describes the object of the evaluation including the results chain, meaning the ‘theory of change’ that underlies the programme being evaluated. This theory of change includes what the programme was meant to achieve and the pathway (chain of results) through which it was expected to achieve this.

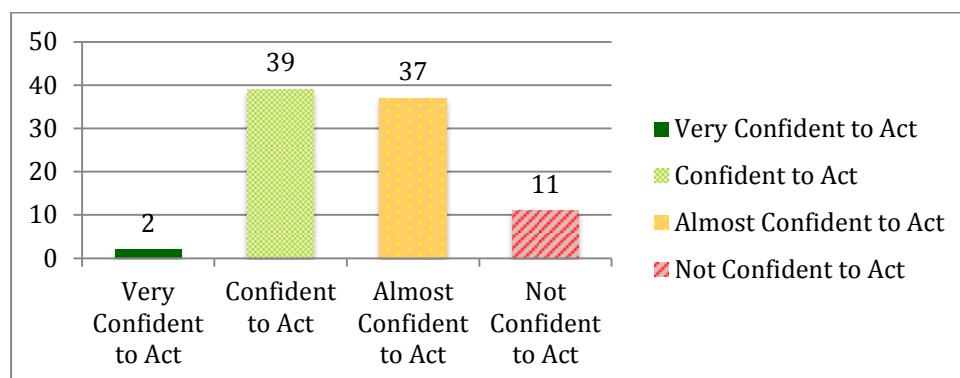
The context of key social, political, economic, demographic, and institutional factors that have a direct bearing on the object is described. For example, the partner government’s strategies and priorities, international, regional or country development goals, strategies and frameworks, the concerned agency’s corporate goals and priorities, as appropriate.”

The object of the evaluation was one of the highest rating sections, with 46% of reports being placed as confident or very confident to act. Of the remaining reports, only 12% were rated as not confident to act, suggesting that a significant number of reports could be brought up to UNICEF standards, largely by having greater clarity in describing the object of the evaluation. The two reports that were rated as outstanding in Section A were the Egypt Forward Looking Assessment of the Community Schools Project, and the Niger Cash Transfer for Protection of Blanket Feeding Final Evaluation.

The two reports rated as *outstanding* both differentiated themselves through demonstrating deeper understanding of the issues involved in the object of the evaluation. Whilst reports rated as *confident* did not explore so deeply, they tended to have a clear description of the object, stakeholder contributions and context. Clarity was emphasised most by reviewers, especially in relation to presenting the theory of change and expected results.

This clarity was not found in reports rated as *almost confident*, which tended not to present the contribution of different stakeholders to the object or described them in broad sweeping terms but without clear information. The 12% of reports rated as *not confident* did not identify the object clearly, explain the current implementation stage, or provide information on how the object was expected to achieve results.

Figure 6: Spread of results for Section A



In general, the distribution of different ratings with Section A is almost exactly the same as the 2009 meta-evaluation. Description of the object, context and current implementation status was strongest, with 57% and 64% of reports rated as satisfactory in these regards. Information about the stakeholders and their contributions was less well covered, with 47% of reports rated satisfactory. The issue of greatest concern remained clarity around the theory of change of the evaluation object, with only 36% of reports rated satisfactory and 33% of reports rated as fully not confident.

Section B: Evaluation purpose, objectives and scope

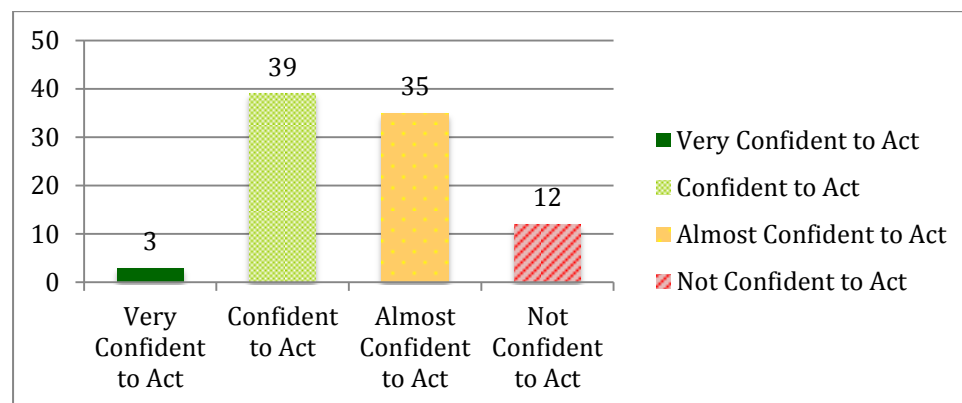
“The purpose of the evaluation is clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed, how the information will be used. The report provides a clear explanation of the evaluation objectives and scope including main evaluation questions and describes and justifies what the evaluation did and did not cover. The report describes and provides an explanation of the chosen evaluation criteria, performance standards, or other criteria used by the evaluators.”

Purpose, objectives and scope was the second strongly rated section, with 47% of reports reaching UNICEF standards. A further 39% of reports were almost satisfactory and could be improved fairly simply: as with Section A, clarity was a big differentiating factor. Three reports were rated as outstanding: Mali’s Evaluation du Programme de Protection de l’Enfant; Pakistan’s Inter-Agency Real Time Evaluation; and Sudan’s Evaluation of the Go to School Initiative.

Reports rated as *outstanding* were differentiated by clearly defined rationale for the selection of evaluation criteria and scope, and discussion around the use of OECD DAC criteria and evaluation questions. Also making reference to OECD DAC criteria,

reports rated as *confident* had clear purpose, objectives, criteria and scope, but lacked the same level of discussion. *Almost confident* reports covered purpose, objectives, criteria and scope, but needed to be clearer around these. *Not confident* reports either did not cover these issues or were confused and disconnected from each other.

Figure 7: Spread of results for Section B



Overall, 62% of evaluation reports were found to have clear purpose, objectives and scope: approximately the same as the previous year (63%). By comparison the use of an appropriate evaluation framework was much weaker – with 59% of reports unsatisfactory – but has demonstrated slight improvement over 2009 reports (65%). In terms of both clarity of purpose and evaluation framework, a large number of reports (almost 30%) could be brought up to standard with a little more work.

Section C: Evaluation methodology

“The report presents a sufficiently detailed description of methodology in which methodological choices are made explicit and justified and in which limitations of methodology applied are included. The report gives the elements to assess the appropriateness of the methodology. Methods as such are not ‘good’ or ‘bad’, they are only so in relation to what one tries to get to know as part of an evaluation. Thus this standard assesses the suitability of the methods selected for the specifics of the evaluation concerned, assessing if the methodology is suitable to the subject matter and the information collected are sufficient to meet the evaluation objectives.”

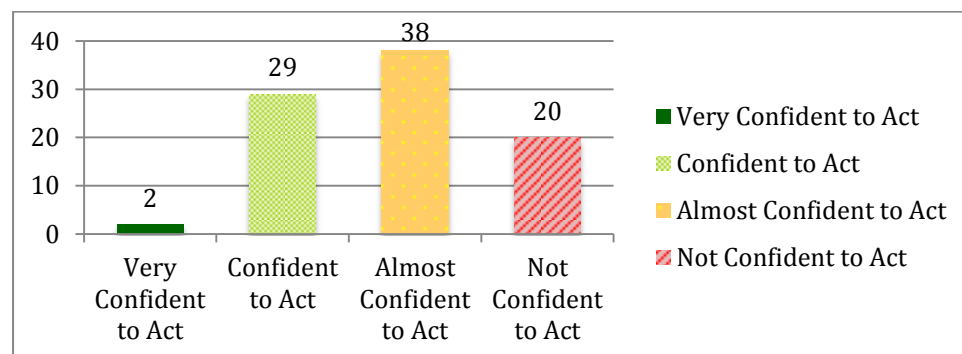
Methodology was generally found to be an area of some concern across all evaluation reports. Indeed, those reports identified as outstanding overall tended to be differentiated largely by having strong methodology sections, because these set up

the rest of the report to deliver convincing evidence. Having stated this, although two-thirds of reports were rated as unsatisfactory, 66% of these could be improved up to UNICEF standards with a little more work.

The two reports rated as having *outstanding* methodology sections were Egypt's Forward Looking Assessment of the Community Schools Project and Kenya's Operational and Impact Evaluation of the Cash Transfer Programme for Orphans and Vulnerable Children. These reports were noted for their transparency about the limitations of the evidence that would be generated. Reports rated as *confident* also tended to discuss the limitations of appropriate methodologies, but many had outstanding issues around human rights, gender, equity and ethics.

The same issues with human rights, gender, equity and ethics affected *almost confident* reports, but these additionally tended to have insufficient descriptions of data limitations. Reports that were rated as *not confident* did not explain data sources, and largely described evaluation tools rather than methodologies.

Figure 8: Spread of results for Section C



Methodological robustness was largely the same as for 2009 reports, with 39% of evaluations found to be satisfactory in this regard (four were deemed to be very confident to act). The level of stakeholder participation (40%) and the quality of data collection (46%) both declined in quality compared to the previous year (51% and 59%). By comparison, the inclusion of results based management showed slight improvement over the previous year; although 27% of reports were still fully not confident in this regard.

Of continued significant concern, however, is the level to which ethics is included in these reports: which has declined further compared to 2009 reports to only 10% satisfactory across both ethics questions. Human Rights, gender and equity also suggested deterioration, but are not comparable due to changes in the rating

questions for these issues introduced for this year's review. These issues are all addressed in more detail in the special section on equity, below.

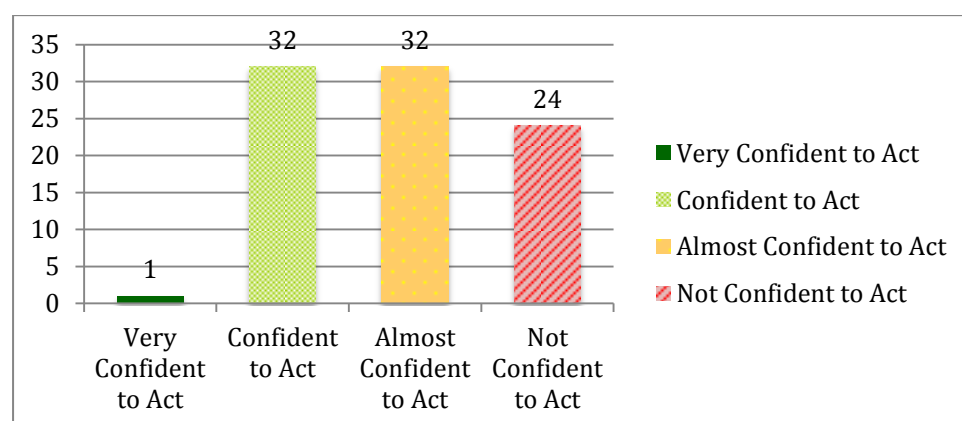
Section D: Findings and conclusions

“Findings respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report. They are based on evidence derived from data collection and analysis methods described in the methodology section of the report.”

Conclusions present reasonable judgments based on findings and substantiated by evidence, providing insights pertinent to the object and purpose of the evaluation.”

The findings and conclusions section showed some considerable concern, and was the weakest section in reports rated as unsatisfactory overall. Over a quarter (27%) of reports were found to be fully not confident and a further 36% required additional work to bring them up to UNICEF standards. Only one report was rated as *outstanding* in regard to its findings and conclusions, Kenya's Operational and Impact Evaluation of the Cash Transfer Programme for Orphans and Vulnerable Children. This report differentiated itself by not only having clear and thorough findings supported by evidence, but also logical conclusions specifically geared to the needs of policy makers.

Figure 9: Spread of results for Section D



Those reports that were rated as *confident* used very clear analysis of data to generate convincing evidence and logically derived conclusions. By comparison, reports rated as *almost confident* also included clear findings and conclusions, but lacked the robust analysis and transparent use of data required to be convincing. Finally,

reports rated as *not confident* tended to be unclear, did not use evidence, presented data as findings, and did not have a logic storyline between findings and conclusions.

Whilst there was a slight improvement in the use of cost analysis (from 21% to 22%), this still means that less than a quarter of reports rate as satisfactory. A majority of reports (55%) did manage to deal satisfactorily with presenting the evaluation object's strengths and weaknesses, and the future implications of these. Reports witnessed slight falls in satisfactory 'contribution, causality and logical conclusions' (from 43% to 40%) and the 'logic of findings' (from 50% to 46%).

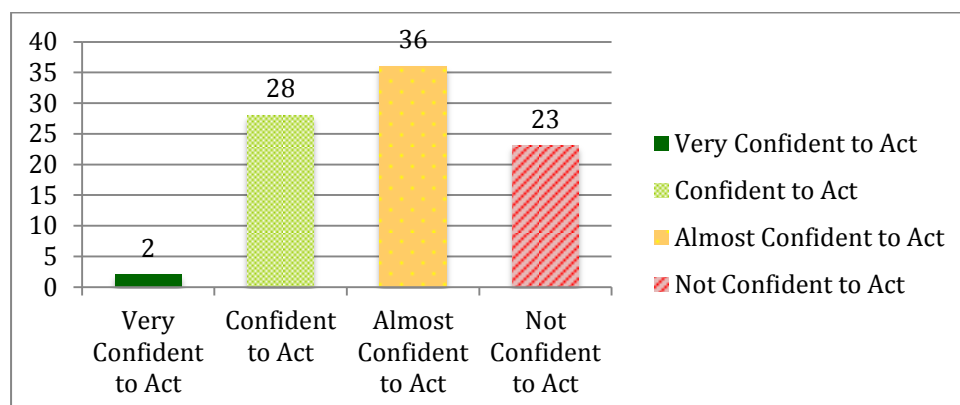
Section E: Recommendations and lessons learned

“Recommendations are relevant to the object and purpose of the evaluation, are supported by evidence and conclusions, and were developed with involvement of relevant stakeholders. Recommendations clearly identify the target group for each recommendation, are clearly stated with priorities for action, are actionable and reflect an understanding of the commissioning organization and potential constraints to follow up”

Section E was the weakest performer mostly because of unsatisfactory lessons learned ratings. Overall, only 34% of evaluations were found to satisfactorily meet UNICEF standards. Nevertheless, there were still two examples of outstanding evaluation reports: Pakistan's Inter-Agency Real Time Evaluation, and Rwanda's Post-Introduction Evaluation of 7-valent Conjugate Pneumococcal Vaccine.

These *outstanding* reports not only provided clear and relevant recommendations, but also identified model-practice for use within and beyond the evaluation object. Those reports rated as *confident* also drew out clear, relevant and actionable recommendations grounded in evidence, but many still had problems in correctly developing lessons learned. The 40% of reports rated as *almost confident* often had large numbers or unclear recommendations, poor linkage to conclusions, and incorrectly interpreted lessons learned. *Not confident* rated reports missed out recommendations; or had untargeted, not specific and non-evidence-based recommendations.

Figure 10: Spread of results for Section E



Despite the poor performance of this section overall, 46% did actually have relevant and clear recommendations (a decline from 52% the previous year). As with the 2009 reports, only 32% of reports were deemed to have recommendations to a satisfactory level of usefulness. The real area of concern, however, relates to the correct identification of useful lessons learned, which declined 11 percentage points year-on-year to a level where 85% of reports do not meet basic UNICEF standards. It would appear that this is still largely due to misunderstanding what lessons learned are; and, where they are present, failing to demonstrate the logical link from the theory of change, through the findings and conclusions to these.

Section F: Report structure, logic and clarity.

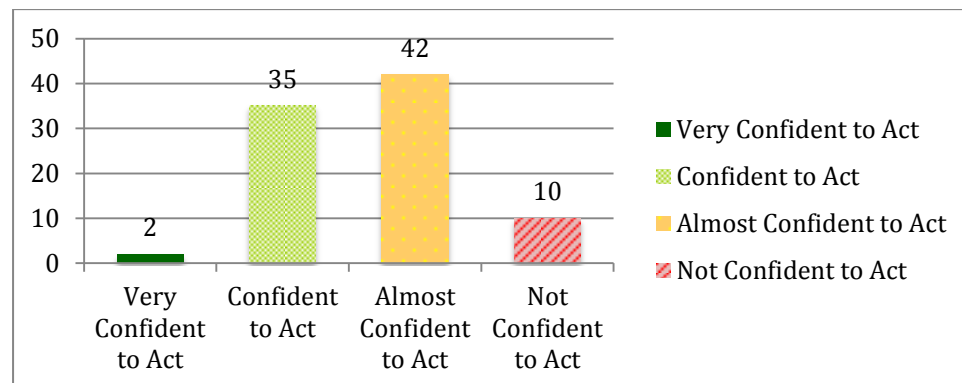
“The report is logically structured with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations). It reads well and is focused.”

One of the stronger sections, 42% of reports rated as satisfactory in relation to the report structure, logic and clarity. It also had the smallest number of reports rated as fully not confident, equating to just over 10%. There were two outstanding reports in terms of structure and executive summaries: Sudan’s Evaluation of the Go to School Initiative, and Albania’s Country Led Evaluation of Delivering as One.

Reviewers noted the concise and well written style of *outstanding* reports, and their subsequent ability to quickly and confidently inform decision making. Those reports rated as *confident* were also clear, logical, made good use of summaries, and had useful executive summaries. By contrast, reports rated *almost confident* also tended to be well structured, but did not have executive summaries that could stand alone and inform decision making. Finally, those reports rated as *not confident* either did

not have executive summaries, or mixed-up findings, conclusions and recommendations sections so as to make them unusable for decision makers.

Figure 11: Spread of results for Section F



The style and presentation of reports demonstrated slight improvement over the previous year (47%), with 50% of evaluations satisfactorily meeting UNICEF standards. Additionally, 53% of reports were deemed to have executive summaries that met UNICEF standards, although this was reduced to 31% on the specific question of usability for decision-making. Executive summaries were not included in 14% of reports, meaning that there were only five reports submitted with executive summaries that were found to be fully not confident. Whilst 86% of reports did have executive summaries attached, this is an indicator – like including the TOR in the annex – where achieving 100% should be feasible simply through introducing basic checks in the contract payment process.

Special Section: Findings on Equity

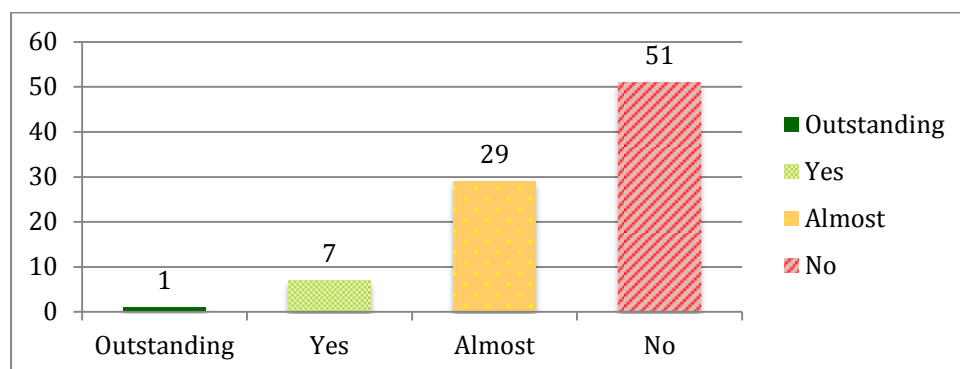
As part of this meta-evaluation, a pilot special analysis of the performance of evaluation reports in relation to equity was requested in order **to generate a baseline** for future reviews. This reflects UNICEF's institutional interest in the national and sub-national distribution of social protection benefits; and the impacts of different global and national economic policies among socio-economic, gender and age groups. The analysis was made possible because of the introduction of a specific equity indicator for this review, based on the adoption of recommendations made in the previous global meta-evaluation.

Overall, only 9% of reports included equity issues to a satisfactory level, according to UNICEF standards: one report was rated outstanding and seven were rated as confident. The outstanding report was *Egypt's Forward Looking Evaluation of the Community Schools Project*, appropriately titled “*Reaching the Unreached*.” The seven *confident* rated reports³ came from Kenya, Montenegro, Myanmar, Pakistan, Corporate, Tanzania, and Albania.

Of the remaining reports, 33% could have reached a satisfactory standard with additional work. Reviewers' comments on these evaluations suggest a trend of equity, gender and human rights issues being discussed in the opening sections of reports; but not carried through the analysis and interpretation of disaggregated data, or used to build evidence about the experiences of different groups.

³ Cash Transfer Programme for Orphans and Vulnerable Children, Kenya: Operational and Impact Evaluation, 2007-2009, Final Report, Kenya; External Evaluation of the Montenegro "Juvenile Justice System Reform Project", Montenegro; Independent Evaluation of UNICEF Education Programme, Improving Access to Quality Basic Education in Myanmar (2006-2010), Myanmar; Inter-Agency Real Time Evaluation of the Humanitarian Response to Pakistan's 2009 Displacement Crisis, Pakistan; Inter-Agency Real Time Evaluation of the Humanitarian Response to Pakistan's 2010 Flood Crisis, Pakistan; A Report on Evaluation of Community Justice Facilitation Project, Tanzania; Country Led Evaluation Delivering As One Albania, Albania.

Figure 12: the inclusion of equity issues in reports



A very strong correlation was found in the coverage of equity issues by reports and their overall quality rating. No report rated overall as below satisfactory received a *confident* rating in relation to equity. Eight reports rated as *not confident* for equity managed to rate as satisfactory overall, whilst nineteen reports rated as *almost confident* for equity were rated as satisfactory or outstanding. All except one of the reports rated not confident overall were also not confident in relation to equity. The report found to be *outstanding* on equity was also found to be outstanding overall.

Regional coverage of equity was mixed. CEE/CIS and ESARO both had two reports rated confident for equity; EAPRO, ROSA, Corporate, and MENA had one satisfactory report each, with the latter being rated outstanding. Neither TACRO nor WCARO had any reports rated as satisfactory for equity: once more raising the question of whether non-English language evaluators have access to the same level or intensity of guidance. WCARO had the highest proportion (over 80%) of reports rated as *not confident*, whilst ESARO had the highest number (14). Corporate evaluations had the highest proportion (60%) of *almost confident* reports.

Proportionately, Real Time Evaluations handled equity better than any other purpose, with half of the four RTEs being rated as satisfactory. Evaluations of pilots had the highest number of confidence ratings for equity (three), but more than half were still rated as not confident. The very small samples of policy and country programme evaluations all failed to deal with equity issues adequately, as did humanitarian evaluations.

Evaluations of 'at scale' objects produced the single outstanding equity report, but also the highest number (14) of not confident ratings for equity. After the two policy evaluations (both rated not confident), evaluations of projects produced the highest percentage (over 70%) of not confident ratings. Programme and humanitarian

evaluations stand out as having half of reports rated as almost confident, suggesting high potential for possible improvements.

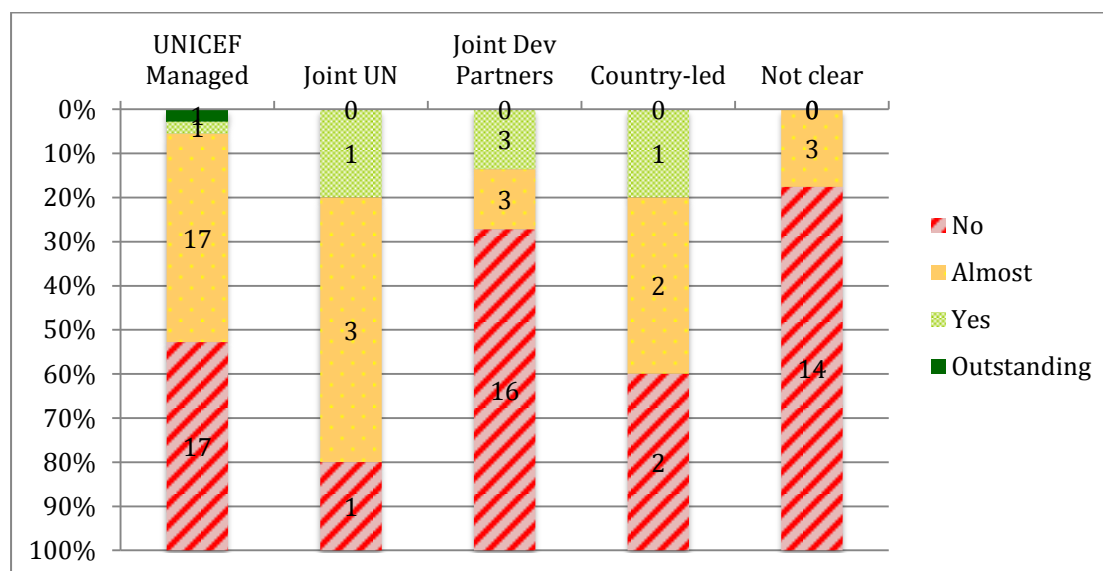
Of the evaluations whose management arrangements could be determined from reports, those conducted jointly with other development partners produced the highest number (three) of satisfactory reports, but almost the highest percentage (over 70%) of not confident ratings. Jointly managed evaluations with UN agencies and country-led evaluations were both small in number (five each), but produced proportionately higher rates of satisfactory reports (both 20%) and lower rates of not confident ratings (20% and 40% respectively).

UNICEF managed evaluations produced one confident report and one outstanding report; but because this management arrangement was by far the highest in number, these represent only 6% of the total. The remaining 34 UNICEF managed reports were found to be unsatisfactory on equity, with half of these being not confident and half being almost confident. This suggests that there is a considerable way to go in terms of UNICEF exerting leadership on equity through the evaluation function.

Finally, we cross-checked coverage of equity against the primary MTSP focus area of reports. Basic Education (MTSP2), Child Protection (MTSP4) and multi-sector evaluation reports were found to be strongest, recording two satisfactory reports each (one of the Basic Education reports was rated outstanding). One satisfactory rating came from HIV/AIDS (MTSP3) reports, but this area also recorded the highest proportion (80%) of not confident ratings (ignoring the single not confident policy and advocacy report). Other areas of concern were Young Child Survival (MTSP1) and Cross-cutting evaluations, neither of which had any satisfactory reports and both of which had above 70% ratings of not confident.

Within the quantitative analysis, no significant pattern was found between report typology and equity. This makes it challenging to determine what drivers exist to advance the cause of equity in evaluations more broadly. The same observation is true for the issues of human rights, gender, and ethics. This would appear to suggest that performance in these areas is largely dependent on individual combinations of evaluators and evaluation managers, rather than an institutional capacity located in some part of the organisation: a major challenge to systemic strengthening.

Figure 13: the coverage of equity by different management arrangements



The differentiating factor for reports that satisfactorily addressed equity was their consistency in mainstreaming equity considerations throughout the evaluation: from design, framework and methodology through to data collection, analysis and conclusions. This allowed the eight reports rated as satisfactory to produce *specific and focused critiques* of the project approach and its shortcomings in addressing equity issues.

One useful approach employed by a report in doing this was found to be conceptualising children as the Key Stakeholder of the evaluation. Other successful reports considered equity in relation to both the implementation and impact of the object, and then extended this to look at the practical implications of these findings.

A total of 29 evaluations attempted to address equity, but were rated as just below satisfactory in doing so: generating a rich seam of information about the challenges of delivering equity-aware evaluation reports. For instance, very few of these reports made reference to UNICEF's policy commitments to equity; and none achieved the consistent thread of equity noted in the satisfactory reports (above).

Whilst equity was found to be addressed less-frequently than gender or human rights, the treatment of issues did not appear to be connected to this: a good human rights report did not necessarily result in a good equity report, and vice versa. In some cases, equity was considered by reviewers to have been addressed indirectly through human rights analysis, but this was not always the rule.

Indeed, the majority of reports do not deal with equity explicitly as a frame of analysis; and when it is this tends to be in relation to elements of the object that are specifically targeted at equity (a similar trend is found in relation to gender). A number of reviews suggested that this could partially be explained by the apparent lack of equity-specific objectives and scope in evaluation TORs. But, this is unlikely to be the whole-story.

In addition to anecdotal evidence of a lack of specific references to equity in evaluation objectives, those reports that do address equity tend to be stronger in discussing the general issue of equity and considerably weaker in specifically analysing equity in the given context. Most reports rated ‘almost confident’ offered consideration of equity as an issue, some few actually collected disaggregated data, and very few managed to generate convincing analysis of this.

It would seem reasonable to infer for this that many evaluators are struggling to understand how to integrate the equity-intention into a robust evaluative process. A couple of reviews also note that this is further compounded by a lack of equity-relevant programme-data available to evaluators even when they do develop analytical frameworks. Furthermore, most of what data is available appears, anecdotally, to be at the output level.

The explanation of ‘lack-of-programme-data’ is insufficient to explain the overall poor performance of evaluation reports; because such a situation should invariable invoke a critique of the object’s monitoring system by the evaluators. This type of assessment of equity in monitoring systems is, however, only present in a very small number of reports. Further making the case for weak knowledge about applying equity among evaluators and evaluation managers is the observation that most evaluation methodologies that *do* attempt to assess equity fail to explain the reasons why the particular disaggregation of groups was chosen.

One final observation that could have relevance to this discussion is the lack of evidence linking analysis of equity and global frameworks. The previous global meta-evaluation proposed that a possible link existed between strong analysis of human rights and the existence of global frameworks such as Child Friendly Schools. The analysis for this years’ report suggests that there is currently no such link in relation to equity. Exploring the dynamics around this further could be a piece of research that UNICEF might wish to consider.

Gender and human rights

Whilst human rights was better covered than equity – with 19% of reports being rated satisfactory – no evaluation was found to be outstanding in this regard. Similarly for gender equality and women’s empowerment issues: 20% of reports were rated as satisfactory, but none were rated outstanding. Reviewer’s feedback provided a similar message to that in equity: that many reports describe gender or human rights concerns, but fail to then translate this language into robust analysis or insight.

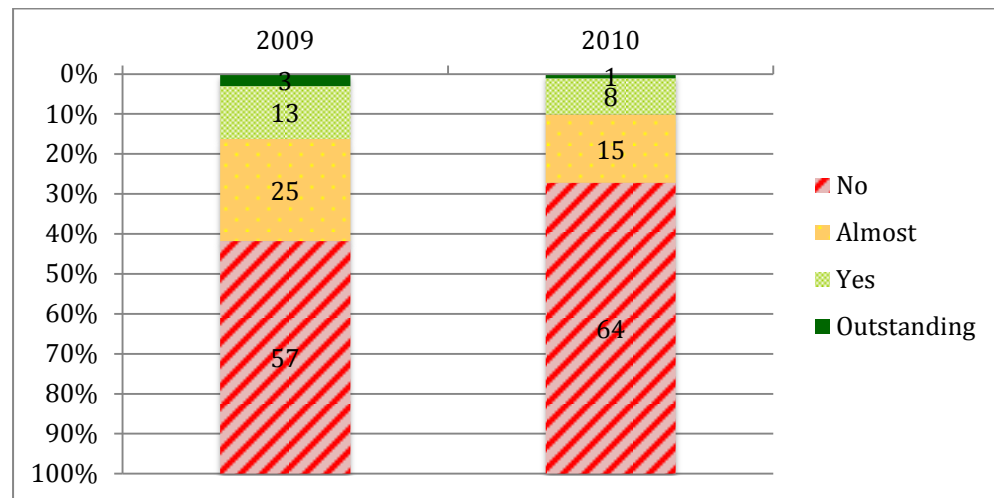
The same trend in equity was observed between reports rating strongly in human rights or gender and rating strongly overall. For both of these issues, however, it was noted that it was possible for reports to do a good job of integrating gender or human rights, and yet rate not confident in totality. This would suggest more of a gap between ‘good evaluators’ and ‘good human rights/gender experts’ than observed purely in relation to equity.

Ethics

The discussion of the ethics issues in evaluations and the explicit introduction of appropriate safeguards was the subject of one of the very first UNICEF technical notes on evaluation quality, yet it remains as the third worst performing area across all evaluations (after equity and lessons learned). Indeed, the level of confidence in ethics sections has actually substantially declined since 2009, from 18% to only 10% of reports reaching the (liberally interpreted) UNICEF standards.

Once again, there are few patterns. Satisfactory evaluations are spread across the regions: MENA produced the most (three), and corporate and WCARO produced the least (none). They are also spread across the MTSP focus areas, although HIV/AIDS worryingly produced no evaluations of satisfactory standard in ethics. Of the nine satisfactory evaluations, one was outstanding in the discussion of ethics (but not its safeguards): Sudan’s Evaluation of the Go to School Initiative. The previous meta-evaluation chose to emphasise core evaluation discipline over individual issues; however, the degree to which ethics is discussed now appears to be on life-support, and warrants renewed highlighting of this concern.

Figure 14: discussion of ethics in 2009 and 2010 reports



Participation

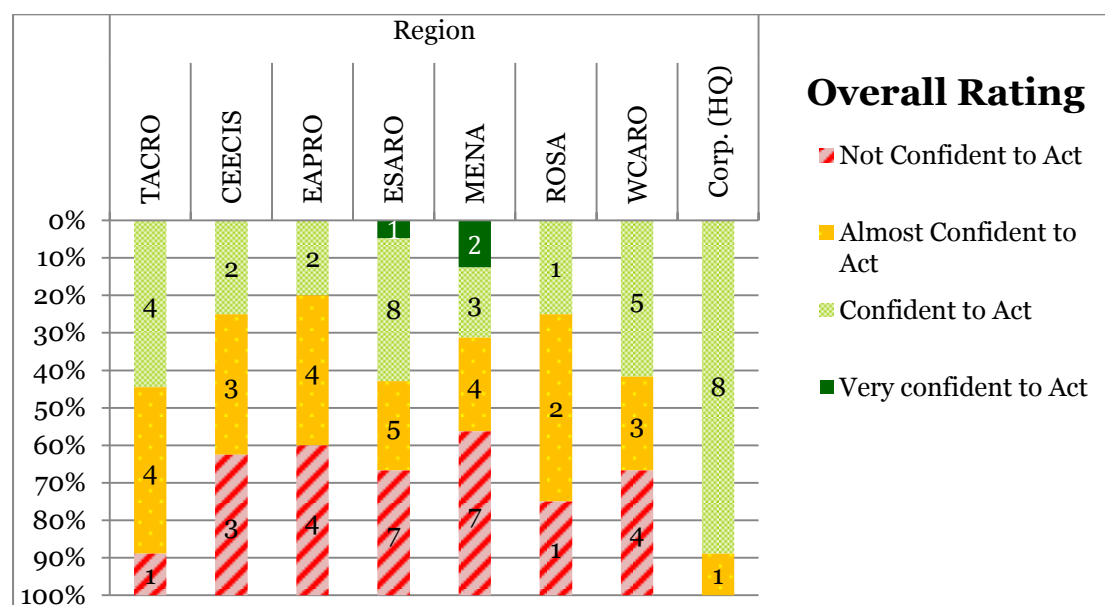
Across all of the evaluations assessed, 45% included a discussion of the level of consultation and participation of stakeholders in the evaluation itself. Of those evaluations that did include this description, half of reports were rated as being satisfactory: meaning that less than a quarter of all reports describe a level of stakeholder participation that meets UNICEF minimum standards. This concern would also seem to be highly relevant context as the organisation attempts to move the equity agenda forwards within its evaluation function.

Regional findings

Regional analysis reveals a mixed performance compared to the previous year. There was slight overall improvement, but a concerning increase in the number and proportion of reports rated as *not confident*, especially in CEE/CIS, EAPRO, MENA and WCARO. The regions that struggled most in the previous meta-evaluation – TACRO, ESARO and WCARO – all demonstrated significant improvement in the proportions of reports rating and satisfactory. MENA and Corporate reports maintained a similar standard to 2009 reports, although the other regions displayed an unfortunate decline in ratings.

Indeed, tracking the regional distribution of performance over the two years 2009-2010 shows that, with the exception of Corporate evaluations, the relative ordering of regions as most-to-least satisfactory reports performed an almost precise reversal. Nevertheless, all except two regions had at least one report where a section was rated outstanding, and both ESARO and MENA both registered their first outstanding overall rated reports.

Figure 15: spread of report ratings by region



Language differences

This meta-evaluation included 16 French language evaluation reports (from ESARO, WCARO and MENA) and six Spanish language evaluation reports (from TACRO). As with the previous meta-evaluation, the overlap between Spanish reports and a single region mean that it is not possible to discount the geographical explanation for

differences, rather than a language explanation (such as having access to less guidance). This is especially true considering that the increase in satisfactory Spanish language reports has accompanied the rise in TACRO's overall performance.

Table 1: proportion of different reports rated confident by language

	Overall	English	French	Spanish
Rated confident	41%	44%	25%	50%

Unlike the previous meta-evaluation, however, French language reports are sourced from a spread of WCARO, ESARO and MENA. They are also greater in number and proportion than in the previous meta-evaluation. With this in mind, there does appear to be a pattern emerging of French language reports rating less well than all of the regional averages from which they came. This would suggest that there could indeed be an argument that having access to lower volumes of mostly translated guidance is adversely affecting the quality of final reports. To be able to assert such a conjecture conclusively, however, is beyond the scope of this methodology.

Examples of interesting and innovative practices

During the reviews, good practices were noted and reported back to evaluation managers. Some of the regional highlights were:

Table 2: examples of good evaluation practice noted in the regions:

Object (Section 1)	☐ N/A
Purpose (Section 2)	☐ Including national partners in the design and execution of the evaluation: and tying this involvement into their other work (MENARO). ☐ Including a table identifying participant involvement and contributions to the evaluation process (TACRO). ☐ Paying attention to the experience with other similar projects in the same country to inform the evaluation (Corporate). ☐ Including a thorough review of the existing legislation and policy documentation (CEE/CIS). ☐ Describing the object of the evaluation, including a thorough description of the context, on a variety of levels (EAPRO).

Methodology (Section 3)

- ☐ Giving care to the engagement with children and the use of UNICEF staff for translation (CEE/CIS).
- ☐ Using a DAC scorecard across several country studies (Corporate).
- ☐ The development of an evaluation matrix in which the DAC
- ☐ Evaluation Criteria are specified within the context of the present evaluation (Corporate).
- ☐ Developing an evaluation matrix of evaluation questions and objectives linked to OECD-DAC criteria (EAPRO).
- ☐ Using detailed evaluations of specific aspects of a pilot intervention (EAPRO).
- ☐ Using as combination of sources of information and ways to collect that information (ESARO).
- ☐ Specifying DAC Evaluation criteria in the evaluation questions.
- ☐ Extensive use of field work and attempts to use a counterfactual (ESARO).
- ☐ A well described participatory approach, combined with the use of a comparison study (WCARO).
- ☐ Consulting stakeholders for the methodological design (WCARO).
- ☐ Considering gender and age as criteria for sampling (WCARO).
- ☐ Thorough field work and use of qualitative and quantitative methods (WCARO).
- ☐ The use and definition of evaluation criteria (WCARO).Emphasising fieldwork in the evaluation, despite the security and time limitations (Corporate).
- ☐ Referencing specific standards and guides for evaluating relevant public health systems and issues (ROSA).
- ☐ Providing a clear definition of each evaluation criterion (ROSA).
- ☐ Monitoring changes in AI knowledge, attitudes and behaviour compared to a counterfactual (MENARO).
- ☐ Using a highly participatory evaluation approach (MENARO).
- ☐ Developing a common evaluation framework to use across countries works well (MENARO).
- ☐ Adapting RCT methodology to meet the needs of UNICEF in terms of ethics (MENARO).
- ☐ Establishing linkages between Result Base Management approach, human rights and the expected results of the object (TACRO).
- ☐ Using a counterfactual to measure programme impact on learners (TACRO).
- ☐ Combining quantitative and qualitative reports to develop a counterfactual (TACRO).

Findings & Conclusions (Section 4)	☐ Relevant use of SWOT (MENARO). ☐ Spending time exploring the issue of attribution (ESARO). ☐ Using quotes from a range of interviewees and including short case studies (ESARO). ☐ Involving the evaluation team in establishing a base line generated objective results post-project (CEE/CIS). ☐ Using robust contribution analysis to provide valid findings without risks of over-claiming (CEE/CIS). ☐ Good use and presentation of first-hand testimony from key stakeholders (CEE/CIS).
Recommendations (Section 5)	☐ The use of provincial and national workshops to validate the findings and conclusions and to develop the recommendations (ROSA). ☐ Using the report to reflect a final consensus shared by stakeholders, developed throughout the evaluation process (MENARO).
Structure (Section 6)	☐ Using annexes to provide further information on key issues included in the report on efficiency and effectiveness (ESARO). ☐ Structuring the evaluation according to a campaign's Theory of Change (Corporate). ☐ Providing regular summaries throughout a complex evaluation (Corporate).

Table 3: examples of good programme practice identified in reports

Corporate	☐ Striking a good balance in the use of resources between building capacity and guidance at HQ and regional levels, and delivery at the country level. ☐ Setting out a process whereby conceptual confusion about adolescents and young people in the human rights frameworks can be overcome. ☐ A co-leadership arrangement for the education cluster, which draws on the strengths of both UNICEF and Save the Children. ☐ Good examples of the use of the cluster system to organise a very large number of agencies. ☐ The use of cash and smart cards in responding to the needs of displaced people.
CEE/CIS	☐ Working in support of Albania's accession process to the European Union seems to have been an effective means to achieve change. ☐ Papa schools: working with fathers on parenting skills. ☐ Focusing on inter-agency linkages and partnerships.
EAPRO	☐ Maintaining a level of sensible expectation within the context of a highly sensitive issue. ☐ The interventions in one country specifically drew on previous work in two others in the region to help in developing a new project suitable for a different context.

ESARO	□ Clear identification of UNICEF's comparative advantage in aligning the project with Government. □ Developing baselines and impact surveys that provide interesting evidence of the unexpected impacts of the programme. □ The contribution of the supplementary feeding programme to other, related, health programmes. □ Supporting a range of NGOs already established and with a track record of working on HIV/AIDS. □ Rigorous pre-planning and planning of a vaccination campaign, including the use of assessments at key stages. □ Taking a novel, community led, approach to dealing with access to justice. □ Attempting to engage with influential traditional authorities on the issue of girls' education.
MENARO	□ One evaluation identifies two important lessons pertaining to immunization programmes in emergency situations: that house-to-house canvassing and vaccination result in better coverage than fixed vaccination posts, and that advocacy and social mobilisation are key to the success of immunization campaigns. □ A systematic attempt to mainstream the community schools approach into the national education system and change the role of UNICEF to a knowledge partner rather than a programme manager. □ Continuing to operate effectively under very difficult security conditions. □ The child-to-child approach appears to offer a very real asset to UNICEF in terms of large-scale impact on school readiness that can be affordably mainstreamed into national systems. □ Drawing on emergency assistance from a WASH expert from Sudan who spoke Arabic resulted in important changes to the latrine technology that was used.
ROSA	□ VA monitoring management by an indigenous NGO that can harmonise with wider public health monitoring. □ Managing a large-scale response across a wide range of organisations. □ Making considerable efforts to coordinate the efforts of different stakeholders and to strengthen the capacity for coordination.
TACRO	□ Using historical knowledge of the political, institutional and social dynamics surrounding the education system in a particular area of Mexico.
WCARO	□ The project made a switch from latrine construction to piloting the Community Led Total Sanitation approach as more appropriate and sustainable in the context. □ Focusing on water utilisation and hand washing in a difficult environment. □ Focusing on both ensuring that the policy environment is right and on developing pilot work across the region.

Conclusions

The following conclusions were developed by analysing the findings for trends in underlying factors that contributed to the performance of evaluation reports. This analysis was grounded in the concept of ‘confidence’, as articulated in the UNICEF Global Evaluation Report Oversight System.

Inconsistency of ratings across time, regions and sectors suggests low institutionalisation of quality evaluation capabilities

This report provided the first opportunity to identify strong patterns in the data that are consistent between 2009 and 2010 reports. What appears to be revealed, however, is that, taken as a whole, no single region outside of corporate evaluations is consistently stronger than any other. For example, if regions are ranked in terms of the average ratings of evaluation reports, the result is almost the polar opposite of the previous meta-evaluation. Similarly, whilst policy evaluations are consistently weak, other MTSP-focus areas also have inconsistent performance.

Where there are potential causal relations, these appear to be complex, multifaceted and not explicable by only one variable. The implication of this is that the eventual quality of a report is largely a result of the unique interactions between the context, evaluation managers, and evaluation team. There is some indication that MTSP areas having longstanding rights-based frameworks – such as Child Friendly Schools under Basic Education and Gender Equality – are more consistently satisfactory than in other areas. However, this is far more tenuous than in the previous meta-evaluation, and MTSP 2 evaluations were still 50% unsatisfactory.

Conjecturing from this, it would appear that – with the exception of corporate evaluations – there are not really dependable ‘stores’ of evaluation reporting capability concentrated anywhere in the system. Great evaluation reports seem to be just as likely to come out from one point in the evaluation system as any other: as do very poor quality reports.

One explanation for this variability could be the methodological limitation of this review: that it assesses *reports* and not evaluations. There does appear to be a few examples where good evaluations appear to be let-down by poorly written reports. However, in general this does not appear to be the case: suggesting that there is merit

in exploring the possibility that evaluative capacity is generally more individualised than institutionalised in UNICEF.

UNICEF appears not to be using its comparative advantage in equity, gender and human rights when it comes to evaluation reporting

As an organisation, UNICEF has been at the forefront of international efforts to recognise human rights, gender equality, women and children's empowerment, and socio-economic equity within the evaluation function. This is not being translated into strong performance in these areas. Our special investigation of equity found it to be the third lowest scoring indicator across the whole assessment.

Explanations for this performance are largely conjecture. For instance, one could be that the renewed focus on equity is still relatively recent for the organisation, and that this has not yet translated into contemporary country-level evaluations. This cannot be the case, however, for human rights, gender or ethics: all of which rate poorly despite being core principles of the organisation. There could also be an explanation that there are insufficiently skilled or affordable independent evaluators in these areas (there was not a single self-evaluation in this year's review). It may be the case, therefore, that having management oversight is simply not enough to consistently deliver satisfactory equity-sensitive evaluations.

The performance of impact evaluations and joint UN evaluations could suggest higher resourcing of evaluations is leading to better reports – but might not

One of the few very strong trends is that impact evaluations rate well on report quality and output evaluations do not: irrespective of MTSP focus area or region. One slightly lesser trend is that evaluations jointly managed with other UN agencies also tended to be stronger overall. On the assumption that impact evaluations and joint evaluations tend to involve more management time and resources than other evaluations, this could imply that higher levels of investment are generally leading to better quality reports – suggesting support for the 'do fewer evaluations, better' line of argument.

It is not an entirely solid conclusion, however, as our analysis suggests that a disproportionate number of impact evaluations use a UNICEF-managed arrangement. This could imply that jointly-managed evaluations are simply delivering well on easier outcome-level questions, whereas UNICEF is independently

tackling more complex and challenging impact-level evaluations that others do not want to (or cannot) get involved in.

Evidence, logic and analysis really counts in delivering quality reports

Our quantitative analysis suggests that outstanding reports are differentiated by the methodology, whereas other reports are most influenced by whether they get the findings-conclusions nexus right. Focusing on the broad detachment between evaluation design and objects' theories of change could help to move from reports full of description to those containing more analysis and logic. Even with the existing methodologies, many more reports could reach a satisfactory standard simply by more clearly developing their evidence base and using it as a grounding for logically argued evaluative conclusions.

Reports tend to score poorly when they leave issues unexplained. For example, various constraints may have tempered the level of participation that was realistically possible. However, by not explicitly including this reasoning the report is likely to rate poor against participation and the appropriateness of the methodology. Furthermore, it may introduce conclusions or recommendations that appear to be inappropriate in light of the type of evidence presented.

In many senses, the GEROS standards are not looking for perfect evaluations, but for transparent and reasoned reports. Comments from reviewers suggest that reports rating unsatisfactory have a tendency to completely miss-out important discussions (ethics, for example), or suddenly introduce findings or conclusions that appear to be grounded only in opinion. An evaluation that has made compromises is thus better served, and will rate higher in terms of confidence, if it has a report that clearly explains what these compromises are, why they came about and what the implications are.

The purpose of Lessons Learned is not understood

To some extent, the inclusion of *lessons learned* within the same review section as *recommendations* acts as camouflage for the underperformance of the former. As a knowledge-centred organisation, it has to be of concern that only 15% of reports adequately demonstrate an understanding or what the purpose and value of lessons learned is. The global evaluation database is potentially an enormous knowledge resource for the spread of good practice in UNICEF; but the current level of

misunderstanding around lessons learned in reports means that most managers would be able to have very little confidence in, or use from, what they find there.

In most cases, reports misinterpret lessons learned as programme or management-specific issues that are only applicable to the object, or as general well-known observations. (A similar issue is that particular constraints facing the evaluation team are often reported as the 'limitations of the methodology'). Were this report to do the equivalent, it might present an insight that was uniquely particular to the relationship between UNICEF and IOD PARC, rather than to UNICEF and its providers in general.

Ethics is an issue on life-support

The previous meta-evaluation purposely avoided including singular issues in the conclusions section, so as to better focus on changes that might bring benefits across multiple indicators of report quality. The decrease the discussion of ethics from an already very low rating, however, warrants a change in this policy. Whilst no conclusions can be drawn regarding how ethical evaluations actually are, the vast majority of evaluation reports simply fail to mention ethics issues.

Indeed, more reports discuss participation of stakeholders (including children) than discuss ethics: an issue that has to be of substantive concern for a child-rights organisation. The GEROS standards are very clear that reports must include a discussion of both ethical issues relating to the object and to the evaluation methodology (including what protection mechanisms were put in place).

Lessons Learned

These lessons learned have been generated through the analysis of the core-evaluation team. They have been adapted based on subsequent dialogue with and contributions from the UNICEF Evaluation Office and comments from regional offices.

Evaluation reports need evaluation skills, but they also need proactive inputs from specialists in rights, gender, ethics and equity.

A consistent finding across 2009 and 2010 is that evaluations undertaken by technical experts in a *research style* most often result in a weaker quality of reports than those undertaken by evaluators. This seems to be the case even where evaluators

are not familiar with the object. It is also the case, however, that most evaluators appear to be unsuccessful in consistently building analytical frameworks that are coherent with human rights, gender, ethics and equity. This suggests a firm need for experts in these areas to be closely involved with evaluation teams – particularly at the initial design stage.

Despite rating poorly in these reports, *equity* could feasibly be an entry point for doing this as it cuts-across economic, social and gender analysis. By contrast, it would appear that transparency and purposefulness about *ethics* is an issue that is best driven by evaluation managers including clear standards in the ToR of what they want to see in the final report.

Conclusions from a single year's meta-evaluation cannot give the whole picture.

Taken in isolation, there are a number of trends apparent in the 2009 and 2010 reports. Several of these, however, do not hold across the two years' reviews. An example is the clear difference in the previous meta-evaluation between the performance of summative and formative reports: an observation that did not hold true this time. The lesson from this experience is two-fold.

Firstly it reinforces the need to exercise caution in interpreting trends from a single meta-evaluation. Secondly, it emphasises the value of having multiple consecutive reviews using comparable methodology. In order to identify the real long-term trends, therefore, the rating matrix used by a meta-evaluation should be as continuous as possible, evolving only in ways that allow comparisons to be made across cohorts. This reinforces the value of qualitative methodologies that can accommodate change whilst maintaining comparable results.

Recommendations

The following recommendations were first generated through the analysis of the core-evaluation team. They have been improved based on subsequent dialogue with, and contributions from, the UNICEF Evaluation Office and regional offices.

Recommendations for UNICEF Evaluation Office

Explore the option of a developing a specific strategy to enhance policy-focused evaluations and policy-relevant evidence in other evaluations

One trend that stands out across both 2009 and 2010 reviews is the under-performance of policy-focused evaluations. Another is the misunderstanding of lessons learned – potentially an important tool in equipping policy advocates. With an organisational commitment to becoming a global knowledge provider, these can be seen as critical business areas for the evaluation function, and may warrant a specific strategy to accelerate their enhancement.

Consider putting communications at the centre of a campaign approach to ethics and equity

As an organisation that has comparative advantage in communications, campaigning and messaging, there would appear to be an opportunity for UNICEF to apply some of this capability internally to influence the behaviour of evaluation managers and suppliers. The two issues that stand out for such treatment are equity and ethics: the deficit in both is substantial, they are specific issues that can be clearly messaged, and both are key tenets of UNICEF's identity.

Consider investigating the real effects of the language gap

Two consecutive meta-evaluations have found reason to infer that non-English language evaluation reports may be hampered by disparities in volume, timeliness and conceptual nativeness of evaluation guidance. This, however, remains conjecture and is also subject to quality reviewing reports in different languages. As it is an issue that appears to be likely to repeat across years, it would seem to be worth decisively exploring the real nature of the effect of language, if any, on evaluation quality.

Recommendations for Regional and Country Offices

Explore strategies for developing a cohort of affordable evaluators specifically trained in the needs of UNICEF evaluation reports

Evaluation report quality appears to be an outcome of a confluence of particular evaluation managers and particular evaluators at a moment in time. As the evaluation manager, UNICEF is ultimately able to exercise substantive influence over the quality of the final report and of the evaluation itself. This makes investment in internal evaluation management capacity appear to be a sensible approach to enhancing and institutionalising performance.

However, UNICEF also has some specific requirements that may not be familiar to many evaluators: particularly in relation to human rights, equity and ethics. Whilst the combination of these skills is available, such blends are often found in companies rather than individual evaluators or ad hoc teams. This has cost implications. A complementary strategy to developing internal capacity would also be to develop long term relationships with affordable local evaluators that would allow them to invest in the specific capabilities required to meet UNICEF's needs.

For instance, UNICEF offices could cooperate to identify evaluators that they use regularly. These evaluators could then be provided with training in UNICEF's specific evaluation needs. Particular attention could be paid to:

- How to tell the 'story' of an evaluation in a UNICEF report, i.e. the logical progression from purpose, through framework and methodology, to findings and conclusions.
- When and how to address equity and human rights in UNICEF evaluations.
- What level of participation UNICEF expects in evaluations, and how to work with the evaluation management team to successfully achieve this.
- How to be explicit in addressing evaluation ethics, both in the process and in the final report.
- What lessons learned are, and best practice for generating prioritised lists of recommendations.
- Writing executive summaries that can stand alone and meet the needs of UNICEF decision makers.

- How the quality of evaluation reports will be assessed, including the GEROS matrix and the concept of *confidence* used by UNICEF. How good ratings can be achieved even in joint evaluations.

Consider developing a pool of experts in human rights, gender, equity and ethics who can advise evaluation teams

A consistent observation across both this and the previous meta-evaluation is the degree of dichotomy that appears to exist between good evaluation reports and strong inclusion of human rights frameworks. In most cases where a standard DAC evaluation framework is used (i.e. reports written by evaluators rather than in the research style of technical specialists), the evaluation criteria are not specifically contextualised to the rights, gender and equity aspects of the evaluated object. Nor is the methodology considered from the perspective of ethics in that context.

To address this issue, UNICEF offices may wish to consider collaborating to develop a pool of technical experts that would be available to advise evaluation teams *at the inception phase* on how to best integrate human rights, gender, equity and ethics into their evaluation framework. Strengthening the framework in this way is likely to have a positive effect on the whole evaluation, including the report. The resource pool could be internal UNICEF experts or an external technical helpdesk facility.

Terms of Reference show progress, but would benefit from clear tables of contents and direct hyperlinks to evaluation standards

Developing this year's review was made significantly easier by the inclusion of many more terms of reference in evaluation reports. This effort must be recognised and dignified (although it still falls short of the 100% that should be possible with a task of this administrative nature). This progress can be built on.

Including a draft table of contents in the terms of reference – with headings such as 'evaluation framework', 'limitations of the methodology', 'ethical considerations', and 'lessons learned' – not only helps to guide the report writers, but can also act as an easy administrative check in signing off the report. It is something that the Evaluation Office does for this report. And, including hyperlinks to the UNICEF/UNEG standards in the appropriate places can enhance it even further:

How to design and Manage Equity-focused Evaluations:

http://mymande.org/sites/default/files/EWP5_Equity_focused_evaluations.pdf

UNICEF Evaluation Report Standards:

http://www.unicef.org/evaluation/files/UNICEF_Evaluation_Report_Standards.pdf

UNICEF Evaluation Criteria:

http://www.unicef.org/evaluation/files/Eval_Criteria_PPPM_2005.pdf

Ethics and your Responsibilities as a Manager:

http://www.unicef.org/evaluation/files/TechNote1_Ethics.pdf

Maximise the value of the inception period by introducing a pro-forma report where evaluators explain how they intend to implement UNICEF guidance, analyse evidence and develop recommendations

Reports rate well that have a clear and transparent methodology (not necessarily spectacular or innovative, just clear), a set of transparently processed evidence (i.e. a presentation of findings rather than data), and a clear set of evaluative conclusions based on this evidence. Many reports that fail to reach UNICEF standards also miss-out important ethical and participatory considerations.

In many instances it is difficult for evaluators to adequately contextualise their methodology when responding to a TOR if they are unfamiliar with the evaluation object. This is one value of including an inception period. One way of boosting the probability that a final report will meet UNICEF standards is to ensure that all necessary issues are covered from the beginning, and are done so in light of UNICEF standards. This will require interpretation of the standards that is specific to each evaluation.

A possible means of maximising the value of any inception period could be requiring each evaluation team to demonstrate that they have adequately understood and taken into consideration UNICEF's basic standards. This could even be assessed through a pro-forma with questions such as:

- ☐ How does the team interpret UNICEF's requirement for equity in this context?
- ☐ How does the team propose to analyse equity in this report?
- ☐ How does the team interpret the UNICEF standard for lessons learned?
- ☐ How will lessons learned be developed, tested and presented in this report?
- ☐ *Etc.*

Annexes

Annex 1: Terms of Reference

1. Background

UNICEF Evaluation Office (EO) put in place an Evaluation Quality Oversight System to monitor the impact of efforts to strengthen the UNICEF evaluation function globally. The system consists of rating evaluation reports commissioned by UNICEF Country Offices (CO), Regional Offices and HQ divisions against the UNEG/UNICEF Evaluation Report Standards. Reports meeting satisfactory rating are made available in the UNICEF Global Evaluation Database (GED).

The database was created on the UNICEF Intranet in 2001. In June 2002, it was made available to the public on the UNICEF Internet website. It currently contains around 3300 records.

UNICEF EO is looking for an institution to ensure the reviewing of and rating the quality of evaluation reports supported by UNICEF country and regional offices all over the world, as well as HQ divisions.

2. Expected results

The selected institutions will review all evaluation reports in English, French and Spanish received by the EO, rate them against UNEG/UNICEF standards, write an executive feedback to be sent to the CO concerned, and make an analysis of trends, key weaknesses and strengths of UNICEF-supported evaluation report.

3. Expected deliverables

Within the Global Evaluation Quality Oversight System, the selected Institution(s) will deliver the following outputs:

A. Executive feedback template

Develop draft executive feedback templates and finalize them based on Evaluation Office comments. The Executive feedback should be one to two pages, and should focus on strengths, weaknesses and recommendations to improve quality of future evaluations reports.

B. Evaluation reports reviewed, rated and executive feedback sent

Review and rate Evaluation reports received in English, French and Spanish against UNEG/UNICEF standards (annex 1) using the UNICEF/UNEG Evaluation Report Review Matrix (annex 2), and send ad-hoc executive feedback as well as the Evaluation Report Review Matrix filled (with a qualitative rating such as “satisfactory”, “good”, etc.) to the concerned UNICEF Office, through the UNICEF Evaluation Office, highlighting strengths, weaknesses and recommendations to improve the quality of future evaluations reports. The completed Evaluation Report Review Matrix with the actual numerical rating should go the Evaluation Office only.

The estimated total number of evaluation reports to be reviewed will be approximately 300, of which about 80% are in English, 15% in French and 5% in Spanish.

Reports must be fully rated and the feedback given within 10 working days of receipt. At times, there may be as many as 20 to be handled within the 10 day period. If

reports to be rated within the 10 working days exceed 20, the rating time will be extended.

C. Global analysis of trends, key weaknesses and strengths of reports reviewed

After reviewing the reports on a yearly basis, produce a Global analysis (no longer than 20 pages) of trends, key weaknesses and strengths of reports reviewed, including lessons learned and good practices on Evaluation reports, and actionable conclusions and recommendations to improve the Evaluation Quality Oversight System as well as the quality of Evaluation reports. The outline of the global analysis will be developed in consultation with UNICEF Evaluation Office. This will enhance COs capacity to ensure future reports meet the criteria and standards as set out by UNICEF/UNEG. As evaluation reports are actively shared internally and externally, it is important that reports shared are of high standards.

4. Management of the system

The Evaluation Quality Oversight System will be managed by the Senior Evaluation Specialist, Systemic Strengthening, with the support of the Knowledge Management (KM) Specialist.

The selected institution(s) will appoint a project manager who will ensure consistency of rating, quality and timely delivery of expected products, and overall coordination with UNICEF Evaluation Office. The project manager will also provide an update on a monthly basis, which will include a tracking matrix highlighting the status of reviews, ratings and executive feedback.

5. Qualifications

- ☐ Excellent and proved knowledge of evaluation methodologies and approaches
- ☐ Proven experience with Quality review of evaluation reports, preferably with UN agencies
- ☐ Proven professional experience in designing and conducting major evaluations
- ☐ Excellent analytical and writing skills in English required. Adequacy in French and Spanish required, with excellence in French and Spanish a strong advantage
- ☐ Familiarity with UNEG/UNICEF evaluation standards is an asset
- ☐ Sectorial knowledge of UNICEF area of intervention (Child survival and development; education; HIV/AIDS; Child protection; Social protection) is an asset
- ☐ Knowledge and expertise of other or similar quality assurance systems will also be an asset

Annex 2: Review Team

The Core Evaluation Quality Team was made up of ten specialist consultants in Monitoring and Evaluation. They were led by **Sadie Watson**, who was responsible for high level quality assurance, **Hatty Dinsmore**, responsible for overall project management, and **Joseph Barnes**, lead author.

Sadie Watson is a Director with over 10 years' experience as a monitoring and evaluation consultant with a specific focus on social development processes and methods. She has a strong background in monitoring, evaluation and organisational reviews and how these processes contribute to organisational learning and improving the way organisations work.

Hatty Dinsmore is a consultant with core skills and experience in project and financial management in international development. Her experience of living and working in East Africa brings further depth and insight into what constitutes effective and efficient project management and brings a significant appreciation for how to build and manage networks of resources.

Joseph Barnes is a senior consultant with a strong grounding in results based management, international development policy, ethics, and disaster reduction. Joseph combines field experience and applied practice in a variety of programming challenges to his consulting work which focuses primarily on organisational analysis and translating strategic programme design into results-based operations.

Annalize Struwig is a principal consultant and has over 20 years' experience as a development consultant and donor advisor specialising in M&E, health systems research and HIV/AIDS. As a consultant, Annalize has worked for a cross section of development agencies, including DFID, the EC, WTO and Irish Aid. Her work for these agencies includes specialist inputs on M&E, as well as the evaluation of projects, sectoral and inter-sectoral programmes and country programmes.

Frank Noij is an associate with 23 years of experience in evaluation and results based management. He has led and participated in evaluations on country programme, thematic, program and project level and has a track record on quality assurance in evaluation. He has worked on organisational capacity development for results based management and supported organisations in adapting their M&E systems to the requirements of a human rights-based approach to development programming.

Julia Betts is a principal consultant specialising in M&E, education and social development. She has over 15 years' experience applying learning and theory to development in practice, working at levels from the grass roots to the international donor agency (DFID). Building on strong analytical and strategic thinking skills Julia's main consulting focus is on evaluation and performance assessment, social analysis and social impact assessment, equity, rights and inclusion (in particular gender analysis, planning and mainstreaming).

Pierre Robert is an associate with over 20 years' experience as an international policy research expert with detailed knowledge of civil society organisations worldwide and strong experience in M&E. Pierre has practical experience of social impact assessments, institutional reform, gender, democratisation, access to justice and social development processes, conflict prevention, and expertise on international organisations, project management and human rights, with a strong background in political and social issues in Asia and Africa.

Ronnie MacPherson is a senior consultant with over 10 years of experience in international development. His work focuses on bringing together organisational planning & development, monitoring & evaluation, information technology, and knowledge management. Ronnie has worked extensively with donors, policymakers, international development NGOs

and UK-based domestic charities, supporting and delivering organisational and strategic development and facilitating collaborative work between diverse organisations. 33

Francis Watkins is a core associate and social development specialist with over 18 years' experience working bilateral and multilateral agencies, governments and international NGOs. Francis has extensive experience in a variety of sectors and has a wide range of experience leading and working in teams in project and programme design, development and implementation of sectoral work, and monitoring and evaluation systems.

Sheelagh O'Reilly is a Director with over 20 years' experience working with international organisations engaged in development and has a wealth of experience in M&E, evaluation management and development effectiveness. Sheelagh has a broad background within the natural resources arena covering industrial management and research, academia and field project implementation linked to policy development. This experience is linked to an understanding of international legal regimes including those associated with environment, human rights, trade (including IPR) and corporate social responsibility.

Marlene Buchy is a core associate with wide-ranging expertise and research in the social dimensions of natural resource management, participatory governance, gender and social impact assessment. She has over 22 years' experience leading research and programmes for bilateral and multilateral agencies across West and Central Africa and South and Southeast Asia.

Martha Nunez is an associate with almost 30 years' experience in project management and development in education, social issues and the social dimension of environmental and conservation initiatives. She has led a wide range of monitoring and evaluation programmes bilateral institutions and NGOs across Latin America where her work has explored the economic and social implications of forestry programmes. Martha has played a key role in shaping environmental policy and strategy which strengthens community development.

Kristen Olsen is a senior consultant with more than 10 years' experience in providing performance assessment for UK and international bodies. Her expertise is in supporting both geographical and learning communities in determining pathways to sustainable development internationally. Her approach brings together a strong background in participatory processes, an academic and professional grounding in the natural resources sector, and a demonstrated expertise in effectively assimilating and communicating knowledge and best practice towards developing robust management, institutions, and planning and delivery processes.

Each consultant brought a depth of evaluation experience in international development and the team was balanced in a wide range of specialisms including; gender, social exclusion, natural resources, child rights, humanitarian development, rural and social development, health, HIV/AIDS, education, environment and human rights. The team included native speakers in English, Spanish and French which ensured high standards of reviewing in each language.

Annex 3: List of Assessed Reports

Impact-level reports are shaded.

Region	Country	Sequence	Title	Overall rating
MENARO	Egypt	2010-004	Reaching the Unreached Forward looking assessment of the 'Community Schools Project' UNICEF - EGYPT	Very Confident to Act
ESARO	Kenya	2010-019	Cash Transfer Programme for Orphans and Vulnerable Children (CT-OVC), Kenya: Operational and Impact Evaluation, 2007-2009	Very Confident to Act
MENARO	Sudan	2010-001	Go To School Evaluation	Very Confident to Act
TACRO	Caribbean Area Office	2010-003	Assessment of Use of H1N1 materials in Dominica & St. Lucia	Confident to act
TACRO	Colombia	2010-001	Evaluación multi-proyecto para la prevención de la vinculación de niños, niñas y adolescentes a grupos armados ilegales y la atención de niños, niñas y adolescentes desvinculados de los grupos armados ilegales	Confident to act
Corporate (HQ)	EO led	2009-005	Children and the 2004 Indian Ocean Tsunami: UNICEF's Response in Indonesia – Country Synthesis Report	Confident to act
Corporate (HQ)	EO led	2010-008	Progress Evaluation of UNICEF in Emergencies and Post-Crisis transition Programmes	Confident to act
ESARO	Ethiopia	2010-001	Mid-Term Evaluation of EU/UNICEF Supported WASH Programme	Confident to act
WCARO	Ghana	2009-014	2nd Performance Monitoring of the IWASH Project	Confident to act
Corporate (HQ)	Global	2010-002	Final Report: Evaluation of UNICEF's Programme and Work in Relation to Adolescents and the Participation of Children and Young People	Confident to act
Corporate (HQ)	Global	2010-003	UNICEF Unite for Children, Unite against AIDS campaign evaluation	Confident to act
Corporate (HQ)	Global	2010-007	Cash Transfer for Protection of Blanket Feeding: UNICEF Emergency Project Niger	Confident to act
Corporate (HQ)	Global/EO led	2010-012	Review of Global Education Cluster Co-leadership Arrangement	Confident to act
Corporate (HQ)	Global/EO led	2010-not allocated yet as yet on the DE	Inter-Agency Real-Time Evaluation in Haiti - 3 Months after the Earthquake	Confident to act
MENARO	Iraq	2010-001	Water Quality Control and Surveillance in Iraq	Confident to act
ESARO	Kenya	2010-005	Evaluation of the Emergency Blanket Supplementary Feeding Programme (BSFT) in five districts of Northern Kenya	Confident to act
WCARO	Liberia	2010-001	Assessing of WASH package interventions in 5 counties of Liberia	Confident to act
ESARO	Madagascar	2010-012	Evaluation on girl-to-girl mentorship strategy	Confident to act
WCARO	Mali	2010-004	Evaluation du Programme Protection de l'Enfant au Mali	Confident to act
TACRO	Mexico	2010-007	Impact and process evaluation of the UNDP-UNESCO-UNICEF and Secretary of Education Joint Programme Construye – T	Confident to act
TACRO	Mexico	2010-008	Design and results evaluation of the education programme CREA Tu espacio	Confident to act
CEE-CIS	Montenegro	2010-003	External Evaluation of the Juvenile Justice System Reform Project	Confident to act
MENARO	Morocco	2010-005	Evaluation Projet Accès Equitable des Jeunes aux NT	Confident to act

Region	Country	Sequence	Title	Overall rating
ESARO	Mozambique	2008-005	ASSESSMENT OF UNICEF NGO – GOVERNMENT PARTNERSHIP PROGRAMME	Confident to act
EAPRO	Myanmar	2010-002	Evaluation of UNICEF Education Programme - Improving Access to Quality Basic Education in Myanmar (2006-2010)	Confident to act
Corporate (HQ)	Pakistan	2010-004	Inter Agency real time evaluation of the humanitarian response to Pakistan's 2009 displacement crisis	Confident to act
ROSA	Pakistan	2010-001	Inter-Agency Real Time Evaluation of the Humanitarian Response to Pakistan's 2010 Flood Crisis	Confident to act
EAPRO	Philippines	2010-001	To assess the results obtained in the frame of CPAP 1999-2003 and 2005-2009, and to capture lessons learnt to inform the new country programme (2012-2016) - [UNICEF Philippines Country Program Evaluation]	Confident to act
ESARO	Rwanda	2010-010	Post-Introduction Evaluation of 7-valent Conjugate Pneumococcal	Confident to act
ESARO	Tanzania	2010-001	Evaluation of Impact of Community Justice Methodology and tools	Confident to act
WCARO	Togo	2010-002	Evaluation of "Allo 111" free helpline for children	Confident to act
WCARO	WCARO Regional	2010-005	Roll Out Evaluation of Community Led Total Sanitation in West & Central Africa	Confident to act
MENARO	Yemen & Bangladesh, China, the Democratic Republic of Congo, Ethiopia, Tajikistan	2010-002	Getting ready for school : A Child-to-Child Approach	Confident to act
ESARO	Zambia	2010-004	Evaluation of Sensitisation of Traditional Leaders on the Importance of Girls' Education	Confident to act
ESARO	Zimbabwe	2010-002	Programme of Support for NAP for OVC Impact /Outcome Assessment	Confident to act
CEECIS	Albania	2010-010	Country Led Evaluation of Delivering as One Albania	Confident to act
TACRO	Bolivia	2010-007	Evaluación del Yuyay Japina	Almost Confident to Act
ESARO	Burundi	2010-002	Assistance aux personnes expulsées de la Tanzanie et appui à la réintégration des personnes rapatriées, expulsées et déplacées (2006-2009) Assistance to persons expelled from Tanzania and support to the reintegration of repatriated, expelled and displaced person (2006-2009)	Almost Confident to Act
EAPRO	China	2010-011	Summary Report and Evaluation: Child Injury Prevention Project (2005-2010)	Almost Confident to Act
TACRO	Colombia	2010-002	Evaluation, consolidation and strengthening of the "I'm a Person, Too" programme for optimum psychosocial development of children under six in the Department of Huila, Colombia - [Evaluación del PROGRAMA "TAMBIÉN SOY PERSONA"]	Almost Confident to Act
MENARO	Djibouti	2010-007	Evaluation of the YCSD programme	Almost Confident to Act
ESARO	Ethiopia	2010-002	Evaluation of Netherlands-UNICEF Water Initiative (NUWI)	Almost Confident to Act
WCARO	Ghana	2009-013	2nd Evaluation of Strategy for Scaling up Community-led Total Sanitation (CLTS) in Ghana	Almost Confident to Act
Corporate (HQ)	Global	2010-005	Response to the Humanitarian Crisis in Haiti following the 12 January 2010 Earthquake. Achievements,	Almost Confident to

Region	Country	Sequence	Title	Overall rating
			Challenges and Lessons to be Learned.	Act
EAPRO	Indonesia	2010-018	UNICEF School and Posyandu Construction Programme in Nangroe Aceh Darussalam (NAD) and Nias, Final Evaluation Report	Almost Confident to Act
MENARO	Iraq	2010-002	Emergency Water Supply to un-served/underserved/ Vulnerable Areas in Baghdad and the IDPs	Almost Confident to Act
TACRO	Jamaica	2010-003	Evaluation of Project "Strengthening Diversion Opportunities, Rehabilitative and Reintegration Programmes for Children Deprived of their Liberty and Not in the Formal School System: An Integrated Livelihood and Remedial Education Project" and Draft Sustainability Plan	Almost Confident to Act
ESARO	Kenya	2010-012	Support to sustainable Water Management and Governance for the Poor in Drought and Flood-Prone Areas in Kenya	Almost Confident to Act
WCARO	Mauritania	2010-009	Evaluation of community-led total sanitation (CLTS)	Almost Confident to Act
TACRO	Mexico	2010-009	Evaluation of the community- based rehabilitation project for disabled children in 4 municipalities of Oaxaca	Almost Confident to Act
ROSA	Nepal	2010-010	Evaluation of UNICEF Nepal Mine Action Activities: Victim-Activated Explosion Injury Surveillance and Mine Risk Education	Almost Confident to Act
EAPRO	North Korea, DPR	2010-002	Evaluation of pilot DeWATS project.	Almost Confident to Act
EAPRO	Papua New Guinea	2010-005	Achieving Development Results Through Partnership: UN Country Programme 2008-2012 and Delivering as One, Papua New Guinea. Independent Mid-term Review, 2010	Almost Confident to Act
WCARO	Senegal	2010-001	Impact du Changement de Normes Sociales sur les Comportements en Milieu Rural au Sénégal	Almost Confident to Act
ROSA	Sri Lanka	2010-002	Evaluation of the Mine Risk Education in Sri Lanka	Almost Confident to Act
MENARO	Sudan	2009-022	Community Radio Listening Groups Project	Almost Confident to Act
MENARO	Yemen	2010-006	Real-Time Evaluation of UNICEF's Response to the Sa'ada Conflict in Northern Yemen	Almost Confident to Act
ESARO	Zambia	2010-005	Evaluation of the Community Life Skills Programme in Central Province for the Period 2007-2009	Almost Confident to Act
ESARO	Zimbabwe	2010-015	Mid-term Evaluation in the 6 EU Water Facility Funded Project	Almost Confident to Act
CEECIS	Albania	2010-002	The Evaluation Report on the National Children's Strategy and Action Plan (2005-2009)	Almost Confident to Act
CEECIS	Ukraine	2010-002	Evaluation of the Ivano-Frankivsk project to improve young child health and development	Almost Confident to Act
CEECIS	Romania	2010-009	Research Report Evaluation of Programmes for MARA (Most-at-Risk Adolescents)	Almost Confident to Act
WCARO	Burkina Faso	2010-003	l'Evaluation de l'appui de l'UNICEF aux OEV et familles vulnérables à travers 7 ONGs, 2006-2010	Not Confident to Act
EAPRO	Cambodia	2010-001	An Assessment of the Law Enforcement against Sexual Abuse, Sexual Exploitation & Trafficking of Children (LEASETC) Project	Not Confident to Act
EAPRO	Cambodia	2010-010	Evaluation of the Accelerated Learning and Multi-Grade Teaching Programmes	Not Confident to Act

Region	Country	Sequence	Title	Overall rating
EAPRO	China	2010-019	Mid-Term Evaluation of Millennium Development Goals Fund (MDGF) China Youth Employment & Migration (YEM)	Not Confident to Act
WCARO	Democratic Republic Congo	2010-010	Etude CAP et Evaluation finale du projet « Appui à l'environnement favorable à l'enfant à travers la participation communautaire en RDC (Kinshasa et Bas-Congo), déc. 2010	Not Confident to Act
MENARO	Djibouti	2010-006	Evaluation of hygiene promotion approaches used by the National Hygiene Committee	Not Confident to Act
MENARO	Djibouti	2010-009	Evaluation of Pre-school programme	Not Confident to Act
MENARO	Egypt	2010-003	Post-Assessment and Evaluation of Avian Influenza Community Education Program in Rural Egypt: Summary Report	Not Confident to Act
WCARO	Gabon	2010-003	Evaluation du fonctionnement du centre de Santé Communautaire d'Owendo	Not Confident to Act
MENARO	Iraq	2010-003, 004, 005, 006	003 Rehabilitation of Sewerage Facilities in select locations in Basrah City 004 Rehabilitation of Water Supply Systems in select locations in Wassit Governorate 005 Extension of Storm Water and Sewer Network in select locations in Kerbala Governorate 006 Rehabilitation/Extension of Storm Water and Sewerage Networks in select locations in Kirkuk City	Not Confident to Act
MENARO	Iraq	2010-007	External Evaluation of joint WHO/UNICEF project "Diseases Eradication, Elimination and Introducing new vaccines (D2-16)"	Not Confident to Act
TACRO	Jamaica	2010-002	Evaluation of the Health and Family Life Education in Jamaica	Not Confident to Act
ESARO	Kenya	2010-013	Evaluation & Review of WESCOORD Structures	Not Confident to Act
ESARO	Kenya	2010-014	Community Health Strategy Evaluation	Not Confident to Act
ESARO	Madagascar	2010-011	Evaluation on the implementation of technical regional staff (ATR)	Not Confident to Act
WCARO	Mauritania	2010-008	Rapport final de l'évaluation du Programme la promotion de l'abandon des MGF mis en oeuvre par le MASEF dans 3 régions à haute prévalence et au Brakna avec l'ONG Tostan	Not Confident to Act
MENARO	Morocco	2010-010	Evaluation du projet de lutte contre le travail des enfants dans le secteur de l'artisanat à Marrakech	Not Confident to Act
ESARO	Mozambique	2010-005	Evaluation of the EU Water Facility Programme (EC-UNICEF Partnership)	Not Confident to Act
EAPRO	Myanmar	2010-001	Evaluation of malaria risk micro-stratification strategy	Not Confident to Act
ESARO	Rwanda	2010-009	Review of the family package project	Not Confident to Act
ESARO	Somalia	2010-003	"Moving Goal Posts" Final evaluation report of Support to Health Sector Development in Somalia	Not Confident to Act
ROSA	Sri Lanka	2010-001	Qualitative Evaluation of Child Friendly Schools in the districts of Badulla and Batticaloa	Not Confident to Act
MENARO	Syria	2010-001	Evaluation of the Child-Friendly Schools Initiative	Not Confident to Act
ESARO	Zimbabwe	2010-011	Summative Evaluation on HIV/AIDS Life Skills Programme	Not Confident to Act
CEECIS	Kazakhstan	2010-008	Final Evaluation: UNICEF component aimed at Strengthening Health and Social Services. Within the 2008-2009 Joint UNFPA/UNICEF/UNDP/UNV Project on Enhancing Human Security in the Former Nuclear Test Site of Semipalatinsk in Kazakhstan	Not Confident to Act
CEECIS	Albania	2010-009	End of Project Evaluation of Child Led Environmental Education	Not Confident to Act
CEECIS	Azerbaijan	2010-002	Azerbaijan Juvenile Justice System Consultation: Evaluation of Rehabilitation Centre and Legal Clinic	Not Confident to Act

Annex 4: Review Tool



UNICEF Global Evaluation Report Oversight System (GEROS) Review Template

Colour Coding	CC	Dark green	Green	Amber	Red	White
	Questions	Outstanding	Yes	Almost	No	Not Applicable
	Section & Overall Rating	Very Confident to Act	Confident to Act	Almost Confident to Act	Not Confident to Act	

The key questions are highlighted as shown here, and are important questions in guiding the analysis of the section
The Cornerstone questions are in column J and are questions that need to be answered for rating and justification of each of the six sections

UNEG Standards for Evaluation in the UN System	UNEG Norms for Evaluation in the UN System	UNICEF Adapted UNEG Evaluation Report Standards
--	--	---

	Response				
Title of the Evaluation Report					
Report sequence number	YYYY/123	Date of Review	DD/MM/YYYY	Year of the Evaluation Report	
Region				Country(ies)	
Type of Report				TORs Present	
Name of reviewer	IOD PARC				
Classification of Evaluation Report					Comments
Geographical (Coverage of the programme being evaluated & generalizability of evaluation findings)					
Management (Managerial control and oversight of evaluation decisions)					
Purpose (Speaks to the overarching goal for conducting the evaluation; its raison d'être)					
Result (Level of changes sought, as defined in RBM: refer to substantial use of highest level reached)					
MTSP Correspondence (Alignment with MTSP focus area priorities: (1) Young child survival and development; (2) Basic education and gender equality; (3) HIV/AIDS and children; (4) Child protection from violence, exploitation and abuse; and (5) Policy advocacy and partnerships for children's rights)					
Level of Independence (Implementation and control of the evaluation activities)					
Timing / Stage					

SECTION A: OBJECT OF THE EVALUATION				
Question	cc	Remarks	A/ Does the report present a clear & full description of the 'object' of the evaluation? The report should describe the object of the evaluation including the results chain, meaning the 'theory of change' that underlies the programme being evaluated. This theory of change includes what the programme was meant to achieve and the pathway (chain of results) through which it was expected to achieve this. The context of key social, political, economic, demographic, and institutional factors that have a direct bearing on the object should be described. For example, the partner government's strategies and priorities, international, regional or country development goals, strategies and frameworks, the concerned agency's corporate goals & priorities, as appropriate. Constructive feedback for future reports Including how to address weaknesses and maintaining good practice	
Object and context				
1 Is the object of the evaluation well described? This needs to include a clear description of the interventions (project, programme, policies, otherwise) to be evaluated including how the designer thought that it would address the problem identified, implementing modalities, other parameters including costs, relative importance in the organization and (number of) people reached.				
2 Is the context explained and related to the object that is to be evaluated? The context includes factors that have a direct bearing on the object of the evaluation: social, political, economic, demographic, institutional. These factors may include strategies, policies, goals, frameworks & priorities at the: international level; national Government level; individual agency level				
3 Does this illuminate findings? The context should ideally be linked to the findings so that it is clear how the wider situation may have influenced the outcomes observed.				

Theory of Change					
4 Is the results chain or logic well articulated? The report should identify how the designers of the evaluated object thought that it would address the problem that they had identified. This can include a results chain or other logic models such as theory of change. It can include inputs, outputs and outcomes, it may also include impacts. The models need to be clearly described and explained.					
Stakeholders and their contributions					
5 Are key stakeholders clearly identified? These include o implementing agency(ies) o development partners o rights holders o primary duty bearers o secondary duty bearers					
6 Are key stakeholders' contributions described? This can involve financial or other contributions and should be specific. If joint program also specify UNICEF contribution, but if basket funding question is not applicable					
7 Are UNICEF contributions described? This can involve financial or other contributions and should be specific					
Implementation Status					
8 Is the implementation status described? This includes the phase of implementation and significant changes that have happened to plans, strategies, performance frameworks, etc that have occurred - including the implications of these changes					
Executive Feedback on Section A Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>					
SECTION B: EVALUATION PURPOSE, OBJECTIVES AND SCOPE					
Question	cc	Remarks			
Purpose, objectives and scope					
9 Is the purpose of the evaluation clear? This includes why the evaluation is needed at this time, who needs the information, what information is needed, how the information will be used.					
10 Are the objectives and scope of the evaluation clear and realistic? This includes: Objectives should be clear and explain what the evaluation is seeking to achieve; Scope should clearly describe and justify what the evaluation will and will not cover; Evaluation questions may optionally be included to add additional details					
11 Do the objective and scope relate to the purpose? The reasons for holding the evaluation at this time in the project cycle (purpose) should link logically with the specific objectives the evaluation seeks to achieve and the boundaries chosen for the evaluation (scope)					
Evaluation framework					
12 Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the Purpose? It is imperative to make the basis of the value judgements used in the evaluation transparent if it is to be understood and convincing. UNEG evaluation standards refer to the OECD/DAC criteria, but other criteria can be used such as Human rights and humanitarian criteria and standards (e.g. SPHERE Standards) but this needs justification. Not all OECD/DAC criteria are relevant to all evaluation objectives and scopes. The TOR may set the criteria to be used, but these should be (re)confirmed by the evaluator. Standard OECD DAC Criteria include: Relevance; Effectiveness; Efficiency; Sustainability; Impact Additional humanitarian criteria include: Coverage; Coordination; Coherence; Protection <i>(This is an extremely important question to UNICEF)</i>					
13 Does the evaluation explain why the evaluation criteria were chosen and/or any standard DAC evaluation criteria (above) rejected? The rationale for using each particular criterion and rejecting any standard OECD-DAC criteria (where they would be applicable) should be explained in the report.					
Executive Feedback on Section B Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>					
SECTION C: EVALUATION METHODOLOGY, GENDER, HUMAN RIGHTS AND EQUITY					
Question	cc	Remarks			
Data collection					
14 Does the report specify data collection methods, analysis methods, sampling methods and benchmarks? This should include the rationale for selecting methods and their limitations based on commonly accepted best practice.					
			B/ Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation? The purpose of the evaluation should be clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed, and how the information will be used. The report should provide a clear explanation of the evaluation objectives and scope including main evaluation questions and describes and justifies what the evaluation did and did not cover. The report should describe and provide an explanation of the chosen evaluation criteria, performance standards, or other criteria used by the evaluators.	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>	
			C/ Is the methodology appropriate and sound? The report should present a transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>	

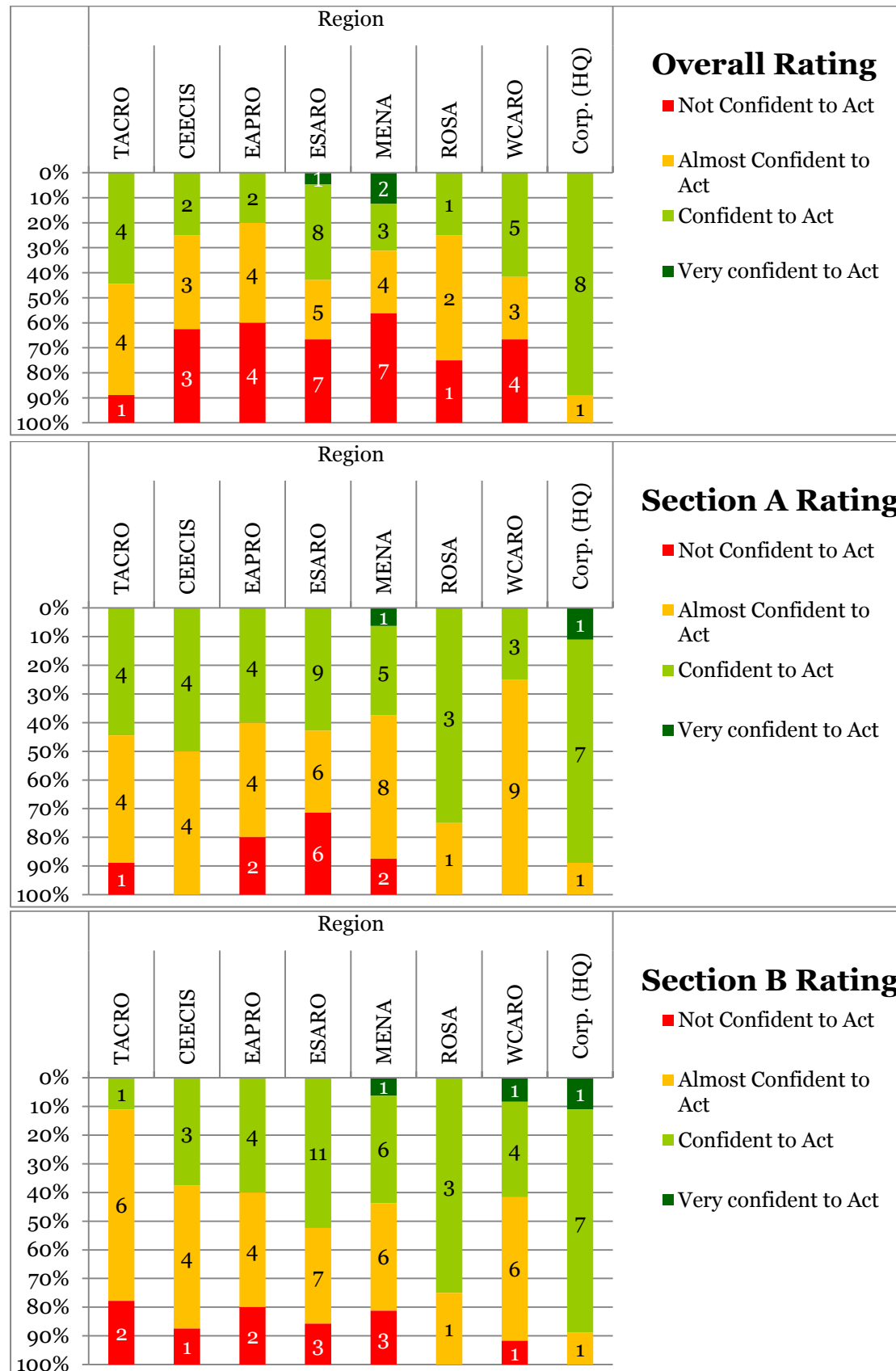
15 Does the report specify data sources, the rationale for their selection, and their limitations? This should include a discussion of how the mix of data sources was used to obtain a diversity of perspectives, ensure accuracy & overcome data limits			specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve the evaluation purposes. The report should also present a sufficiently detailed description of methodology in which methodological choices are made explicit and justified and in which limitations of methodology applied are included. The report should give the elements to assess the appropriateness of the methodology. Methods as such are not 'good' or 'bad', they are only so in relation to what one tries to get to know as part of an evaluation. Thus this standard assesses the suitability of the methods selected for the specifics of the evaluation concerned, assessing if the methodology is suitable to the subject matter and the information collected are sufficient to meet the evaluation objectives.	
Ethics				
16 Are ethical issues and considerations described? The design of the evaluation should contemplate: How ethical the initial design of the programme was; The balance of costs and benefits to participants (including possible negative impact) in the programme and in the evaluation; The ethics of who is included and excluded in the evaluation and how this is done				
17 Does the report refer to ethical safeguards appropriate for the issues described? When the topic of an evaluation is contentious, there is a heightened need to protect those participating. These should be guided by the UNICEF Evaluation Office Technical Note and include: protection of confidentiality; protection of rights; protection of dignity and welfare of people (especially children); Informed consent; Feedback to participants; Mechanisms for shaping the behaviour of evaluators and data collectors				
Results Based Management				
18 Is the capability and robustness of the evaluated object's monitoring system adequately assessed? The evaluation should consider the details and overall functioning of the management system in relation to results: from the M&E system design, through individual tools, to the use of data in management decision making.				
19 Does the evaluation make appropriate use of the M&E framework of the evaluated object? In addition to articulating the logic model (results chain) used by the programme, the evaluation should make use of the object's logframe or other results framework to guide the assessment. The results framework indicates how the programme design team expected to assess effectiveness, and it forms the guiding structure for the management of implementation.				
Human Rights, Gender and Equity				
20 Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity? This could be done in a variety of ways including: use of a rights-based framework, use of CRC, CCC, CEDAW and other rights related benchmarks, analysis of right holders and duty bearers and focus on aspects of equity, social exclusion and gender. Style includes: using human-rights language; gender-sensitive and child-sensitive writing; disaggregating data by gender, age and disability groups; disaggregating data by socially excluded groups				
21 Does the evaluation assess the extent to which the implementation of the evaluated object was monitored through human rights (inc. gender, equity & child rights) frameworks? UNICEF commits to go beyond monitoring the achievement of desirable outcomes, and to ensure that these are achieved through morally acceptable processes. The evaluation should consider whether the programme was managed and adjusted according to human rights and gender monitoring of processes.				
22 Do the methodology, analytical framework, findings, conclusions, recommendations & lessons provide appropriate information on HUMAN RIGHTS (inc. women & child rights)? The inclusion of human rights frameworks in the evaluation methodology should continue to cascade down the evaluation report and be obvious in the data analysis, findings, conclusions, any recommendations and any lessons learned. If identified in the scope the methodology should be capable of assessing the level of identification of the human rights claims of rights-holders and the corresponding human rights obligations of duty-bearers, as well as the immediate underlying & structural causes of the non realisation of rights.; Capacity development of rights-holders to claim rights, and duty-bearers to fulfil obligations.				
23 Do the methodology, analytical framework, findings, conclusions, recommendations & lessons provide appropriate information on GENDER EQUALITY AND WOMEN'S EMPOWERMENT? The inclusion of gender equality frameworks in the evaluation methodology should continue to cascade down the evaluation report and be obvious in the data analysis, findings, conclusions, any recommendations and any lessons learned. If identified in the scope the methodology should be capable of assessing the immediate underlying & structural causes of social exclusion; and capacity development of women to claim rights, and duty-bearers to fulfil their equality obligations.				
24 Do the methodology, analytical framework, findings, conclusions, recommendations & lessons provide appropriate information on EQUITY? The inclusion of equity considerations in the evaluation methodology should continue to cascade down the evaluation report and be obvious in the data analysis, findings, conclusions, any recommendations and any lessons learned. If identified in the scope the methodology should be capable of assessing the capacity development of rights-holders to claim rights, and duty-bearers to fulfil obligations & aspects of equity.				

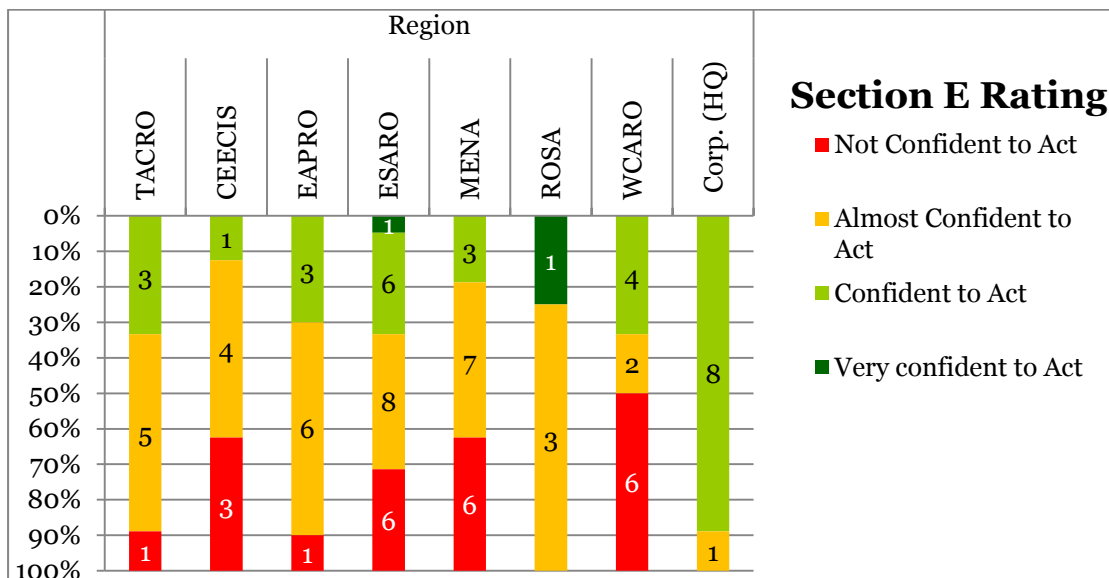
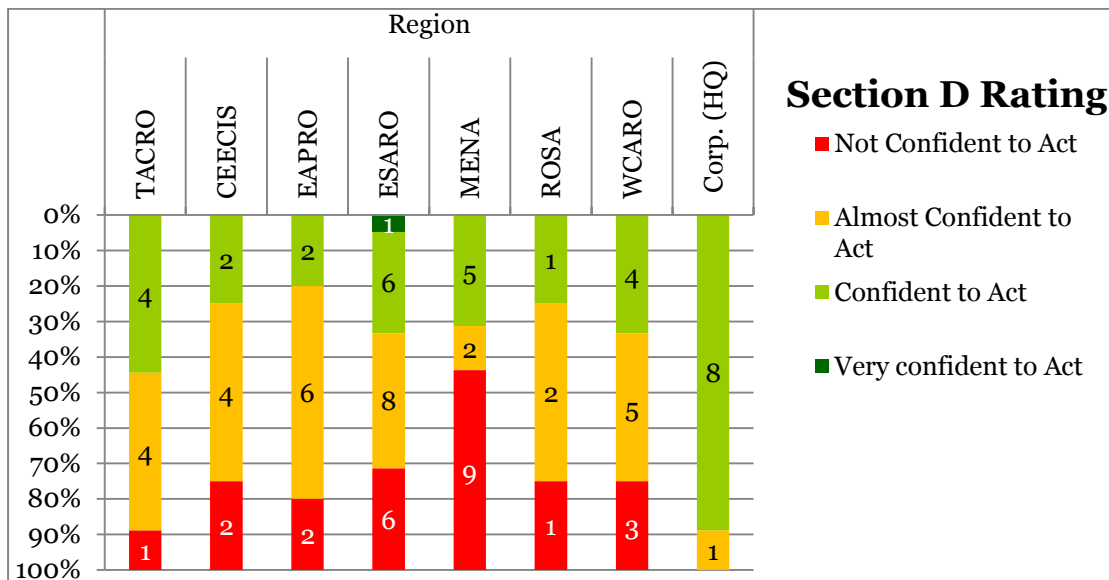
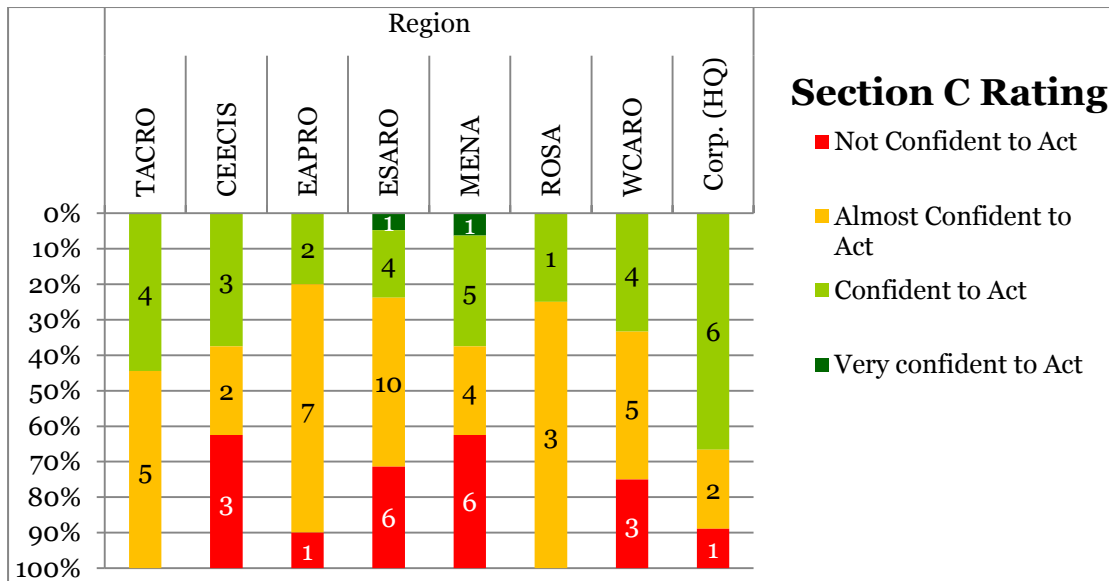
Stakeholder participation					
25 Are the levels and activities of stakeholder consultation described? This goes beyond just using stakeholders as sources of information and includes the degree of participation in the evaluation itself. The report should include the rationale for selecting this level of participation. Roles for participation might include: o Liaison o Technical advisory o Observer o Active decision making The reviewer should look for the soundness of the description and rationale for the degree of participation rather than the level of participation itself.					
26 Are the levels of participation appropriate for the task in hand? The breadth & degree of stakeholder participation feasible in evaluation activities will depend partly on the kind of participation achieved in the evaluated object. The reviewer should note here whether a higher degree of participation may have been feasible & preferable.					
Methodological robustness					
27 Is there an attempt to construct a counterfactual? The counterfactual can be constructed in several ways which can be more or less rigorous. It can be done by contacting eligible beneficiaries that were not reached by the programme, or a theoretical counterfactual based on historical trends, or it can also be a comparison group.					
28 Can the methodology answer the evaluation questions in the context of the evaluation? The methodology should link back to the Purpose and be capable of providing answers to the evaluation questions.					
29 Are methodological limitations acceptable for the task in hand? Limitations must be specifically recognised and appropriate efforts taken to control bias. This includes the use of triangulation, and the use of robust data collection tools (interview protocols, observation tools etc). Bias limitations can be addressed in three main areas: Bias inherent in the sources of data; Bias introduced through the methods of data collection; Bias that colours the interpretation of findings					
Executive Feedback on Section C Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>					
SECTION D: FINDINGS AND CONCLUSIONS					
Question	cc	Remarks	D/ Are the findings and conclusions, clearly presented, relevant and based on evidence & sound analysis?		
Completeness and logic of findings					
30 Are findings clearly presented and based on the objective use of the reported evidence? Findings regarding the inputs for the completion of activities or process achievements should be distinguished clearly from results. Findings on results should clearly distinguish outputs, outcomes and impacts (where appropriate). Findings must demonstrate full marshalling and objective use of the evidence generated by the evaluation data collection. Findings should also tell the 'whole story' of the evidence and avoid bias.			Findings should respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report. They should be based on evidence derived from data collection and analysis methods described in the methodology section of the report. Conclusions should present reasonable judgments based on findings and substantiated by evidence, providing insights pertinent to the object and purpose of the evaluation.	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>	
31 Do the findings address all of the evaluation's stated criteria and questions? The findings should seek to systematically address all of the evaluation questions according to the evaluation framework articulated in the report.					
32 Do findings demonstrate the progression to results based on the evidence reported? There should be a logical chain developed by the findings, which shows the progression (or lack of) from implementation to results.					
33 Are gaps and limitations discussed? The data may be inadequate to answer all the evaluation questions as satisfactorily as intended. In this case the limitations should be clearly presented and discussed. Caveats should be included to guide the reader on how to interpret the findings. Any gaps in the programme or unintended effects should also be addressed.					
34 Are unexpected findings discussed? If the data reveals (or suggests) unusual or unexpected issues, these should be highlighted and discussed in terms of their implications.					
Cost Analysis					
35 Is a cost analysis presented that is well grounded in the findings reported? Cost analysis is not always feasible or appropriate. If this is the case then the reasons should be explained. Otherwise the evaluation should use an appropriate scope and methodology of cost analysis to answer the following questions: o How programme costs compare to other similar programmes or standards o Most efficient way to get expected results o Cost implications of scaling up or down o Cost implications for replicating in a different context o Is the programme worth doing from a cost perspective o Costs and the sustainability of the programme.					
Contribution and causality					
36 Does the evaluation make a fair and reasonable attempt to assign contribution for results to identified stakeholders? For results attributed to the programme, the result should be mapped as accurately as possible to the inputs of different stakeholders.					
37 Do conclusions take due account of the views of a diverse cross-section of stakeholders? As well as being logically derived from findings, conclusions should seek to represent the range of views encountered in the evaluation, and not simply reflect the bias of the individual evaluator. Carrying these diverse views through to the presentation of conclusions (considered here) is only possible if the methodology has gathered and analysed information from a broad range of stakeholders.					

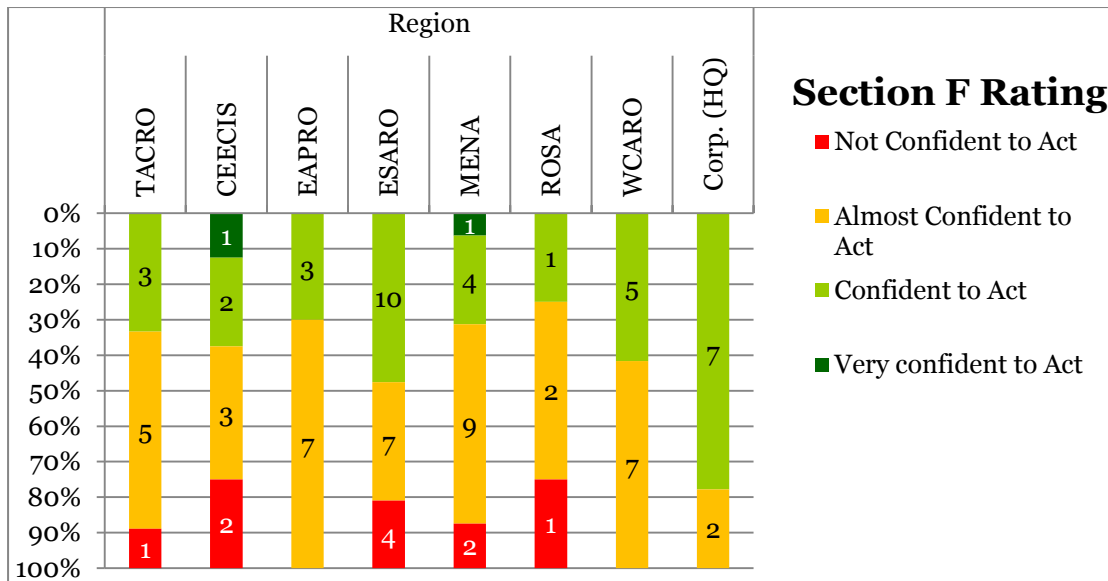
38 Are causal reasons for accomplishments and failures identified as much as possible? These should be concise and usable. They should be based on the evidence and be theoretically robust. <i>(This is an extremely important question to UNICEF)</i>					
Strengths, weaknesses and implications					
39 Are the future implications of continuing constraints discussed? The implications can be, for example, in terms of the cost of the programme, ability to deliver results, reputational risk, and breach of human rights obligations.					
40 Do the conclusions present both the strengths and weaknesses of the evaluated object? Conclusions should give a balanced view of both the stronger aspects and weaker aspects of the evaluated object with reference to the evaluation criteria and human rights based approach.					
Completeness and insight of conclusions					
41 Do the conclusions represent actual insights into important issues that add value to the findings? Conclusions should go beyond findings and identify important underlying problems and/or priority issues. Simple conclusions that are already well known do not add value and should be avoided.					
42 Are the conclusions pitched at a level that is relevant to the end users of the evaluation? Conclusions should speak to the evaluation participants, stakeholders and users. These may cover a wide range of groups and conclusions should thus be stated clearly and accessibly: adding value and understanding to the report (for example, some stakeholders may not understand the methodology or findings, but the conclusions should clarify what these findings mean to them in the context of the programme).					
Executive Feedback on Section D Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>					
SECTION E: RECOMMENDATIONS AND LESSONS LEARNED					
Question	cc	Remarks			
Relevance and clarity of recommendations					
43 Are the recommendations well-grounded in the evidence and conclusions reported? Recommendations should be logically based in findings and conclusions of the report.			E/ Are the recommendations learned relevant and actionable? Recommendations should be relevant and actionable to the object and purpose of the evaluation, be supported by evidence and conclusions, and be developed with involvement of relevant stakeholders. Recommendations should clearly identify the target group for each recommendation, be clearly stated with priorities for action, be actionable and reflect an understanding of the commissioning organization and potential constraints to follow up.	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>	
44 Are recommendations relevant to the object and the purpose of the evaluation? Recommendations should be relevant to the evaluated object					
45 Are recommendations clearly stated and prioritised? If the recommendations are few in number (up to 5) then this can also be considered to be prioritised. Recommendations that are over-specific or represent a long list of items are not of as much value to managers. Where there is a long list of recommendations, the most important should be ordered in priority.					
Usefulness of recommendations					
46 Does each recommendation clearly identify the target group for action? Recommendations should provide clear and relevant suggestions for action linked to the stakeholders who might put that recommendation into action. This ensures that the evaluators have a good understanding of the programme dynamics and that recommendations are realistic.					
47 Are the recommendations realistic in the context of the evaluation? This includes: o an understanding of the commissioning organisation o awareness of the implementation constraints o an understanding of the follow-up processes					
48 Does the report describe the process followed in developing the recommendations? The preparation of recommendations needs to suit the evaluation process. Participation by stakeholders in the development of recommendations is strongly encouraged to increase ownership and utility.					
Appropriate lessons learned					
49 Are lessons learned correctly identified? Lessons learned are contributions to general knowledge. They may refine or add to commonly accepted understanding, but should not be merely a repetition of common knowledge. Findings and conclusions specific to the evaluated object are not lessons learned.					
50 Are lessons learned generalised to indicate what wider relevance they may have? Correctly identified lessons learned should include an analysis of how they can be applied to contexts and situations outside of the evaluated object.					
Executive Feedback on Section E Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>					

SECTION F: REPORT IS WELL STRUCTURED, LOGIC AND CLEAR				
Question	cc	Remarks		
Style and presentation				
51. Do the opening pages contain all the basic elements? Basic elements include all of: Name of the evaluated object; Timeframe of the evaluation and date of the report; Locations of the evaluated object; Names and/or organisations of evaluators; Name of the organisation commissioning the evaluation; Table of contents including tables, graphs, figures and annex; List of acronyms			F/ Overall, do all these elements come together in a well structured, logical, clear and complete report? The report should be logically structured with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations). It should read well and be focused.	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>
52. Is the report logically structured? Context, purpose, methodology and findings logically structured. Findings would normally come before conclusions, recommendations & lessons learnt				
53. Do the annexes contain appropriate elements? Appropriate elements may include: ToRs; List of interviewees and site visits; List of documentary evidence; Details on methodology; Data collection instruments; Information about the evaluators; Copy of the evaluation matrix; Copy of the Results chain. Where they add value to the report				
54. Do the annexes increase the usefulness and credibility of the report?				
Executive Summary				
55. Is an executive summary included as part of the report? If the answer is No, question 56 to 58 should be N/A				
56. Does the executive summary contain all the necessary elements? Necessary elements include all of: Overview of the evaluated object; Evaluation objectives and intended audience; Evaluation methodology; Most important findings and conclusions; Main recommendations				
57. Can the executive summary stand alone? It should not require reference to the rest of the report documents and should not introduce new information or arguments				
58. Can the executive summary inform decision making? It should be short (ideally 2-3 pages), and increase the utility for decision makers by highlight key priorities.				
Executive Feedback on Section F Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>				
Additional Information				
Question	Remarks			
i/ Does the evaluation successfully address the Terms of Reference? If the report does not include a TOR then a recommendation should be given to ensure that all evaluations include the TOR in the future. Some evaluations may be flawed because the TORs are inappropriate, too little time etc. Or, they may succeed despite inadequate TORs. This should be noted under vii in the next section	Problems with the TOR may be noted under Overall Rating Question 3			
ii/ Identify aspects of good practice of the evaluation In terms of evaluation				
iii/ Identify aspects of good practice of the evaluation In terms of programmatic, sector specific, thematic expertise				
OVERALL RATING				
Question	cc	Remarks	OVERALL RATING Informed by the answers above, apply the reasonable person test to answer the following question: Is this a credible report that addresses the evaluation purpose and objectives based on evidence, and that can therefore be used with confidence? This question should be considered from the perspective of UNICEF strategic management.	
i/ To what extent does each of the six sections of the evaluation provide sufficient credibility to give the reasonable person confidence to act? Taken on their own, could a reasonable person have confidence in each of the five core evaluation elements separately? It is particularly important to consider: o Is the report methodologically appropriate? o Is the evidence sufficient, robust and authoritative? o Do the analysis, findings, conclusions and recommendations hold together?				
ii/ To what extent do the six sections hold together in a logically consistent way that provides common threads throughout the report? The report should hold together not just as individually appropriately elements, but as a consistent and logical 'whole'.				
iii/ Are there any reasons of note that might explain the overall performance or particular aspects of this evaluation report? This is a chance to note mitigating factors and/or crucial issues apparent in the review of the report.	ToRs Other			
Executive Feedback on Overall Rating Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>				

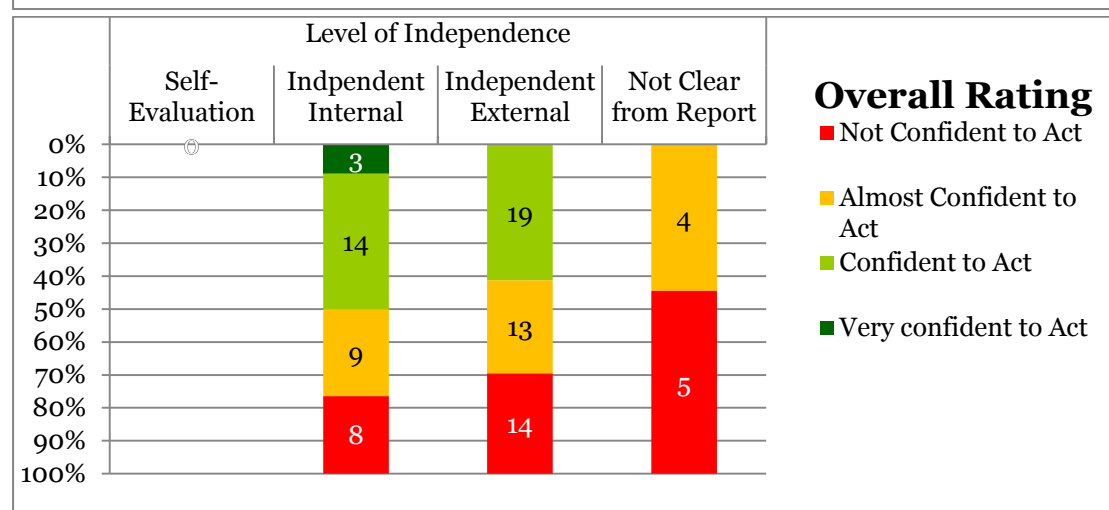
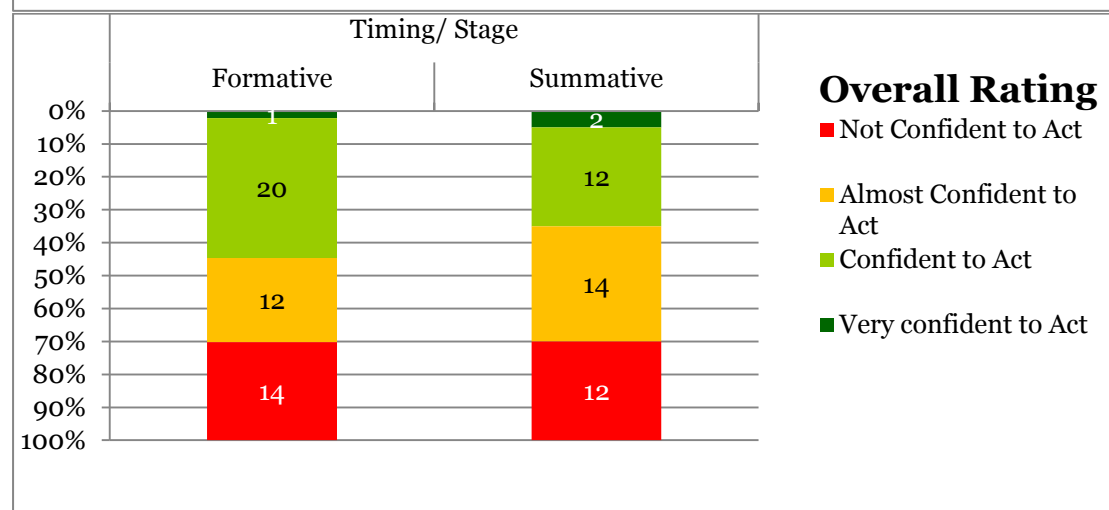
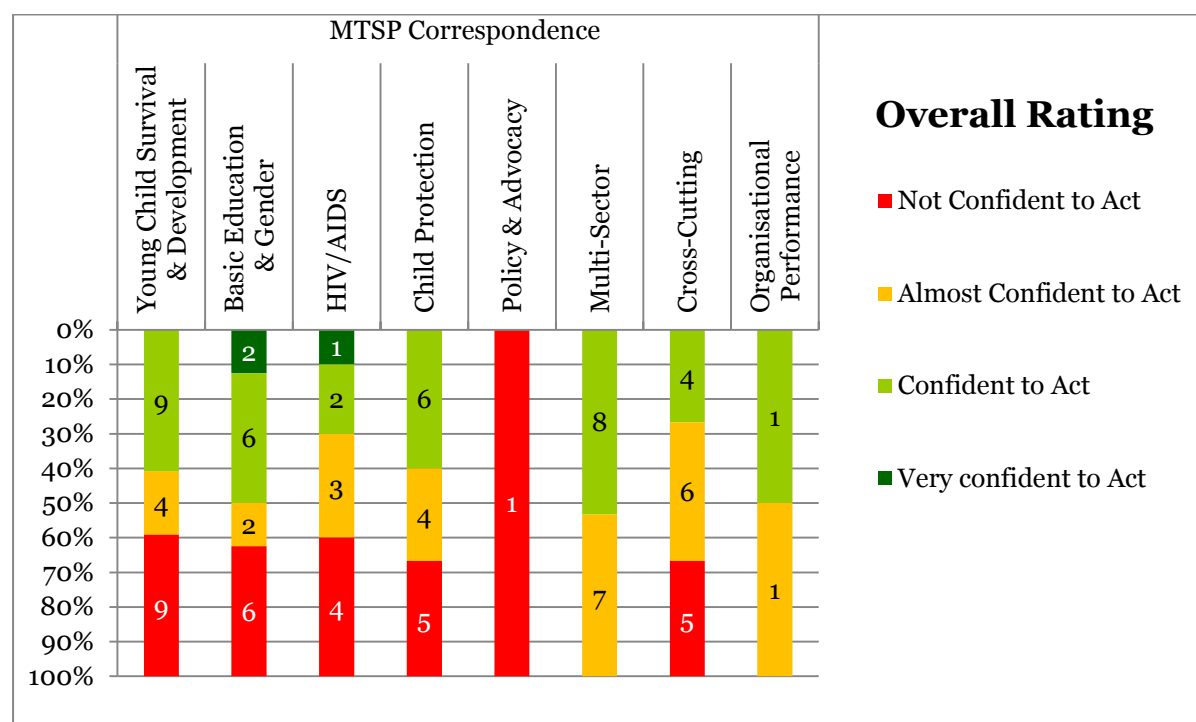
Annex 5: Regional Ratings Graphs

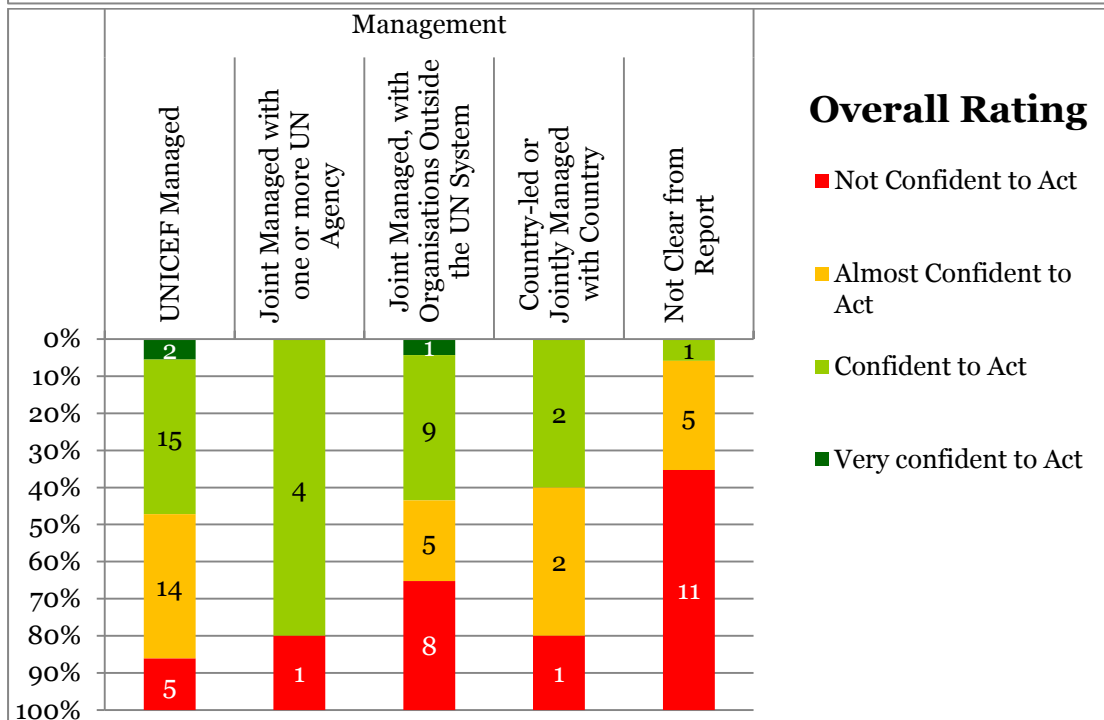
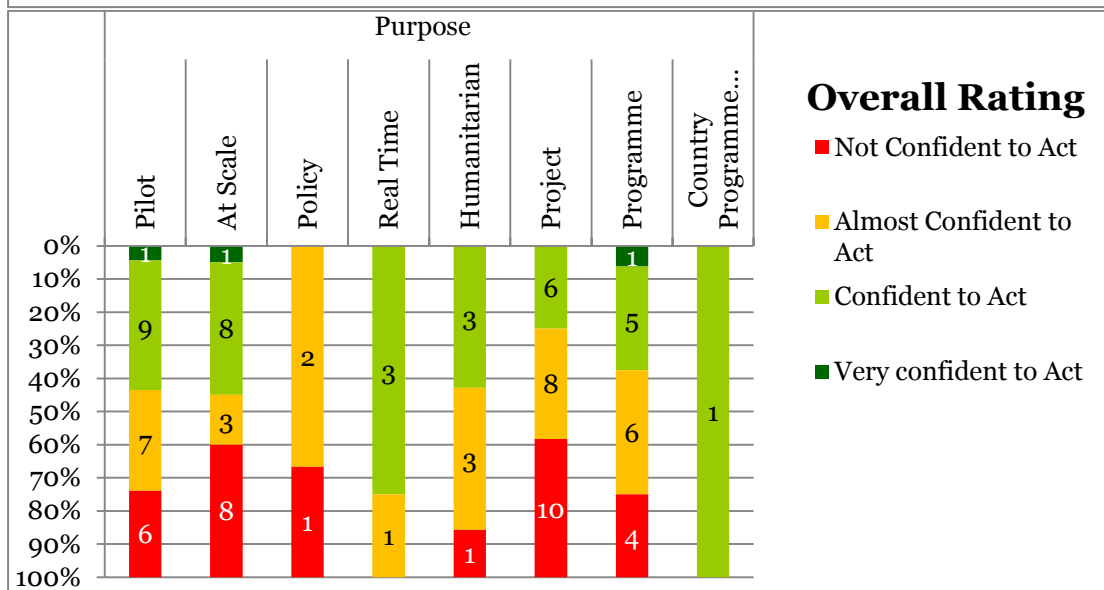
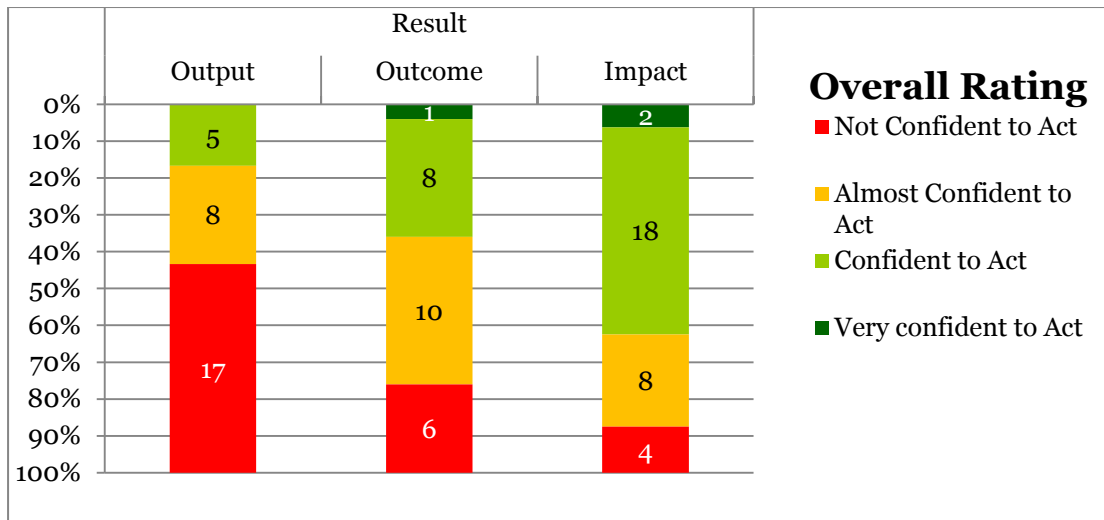


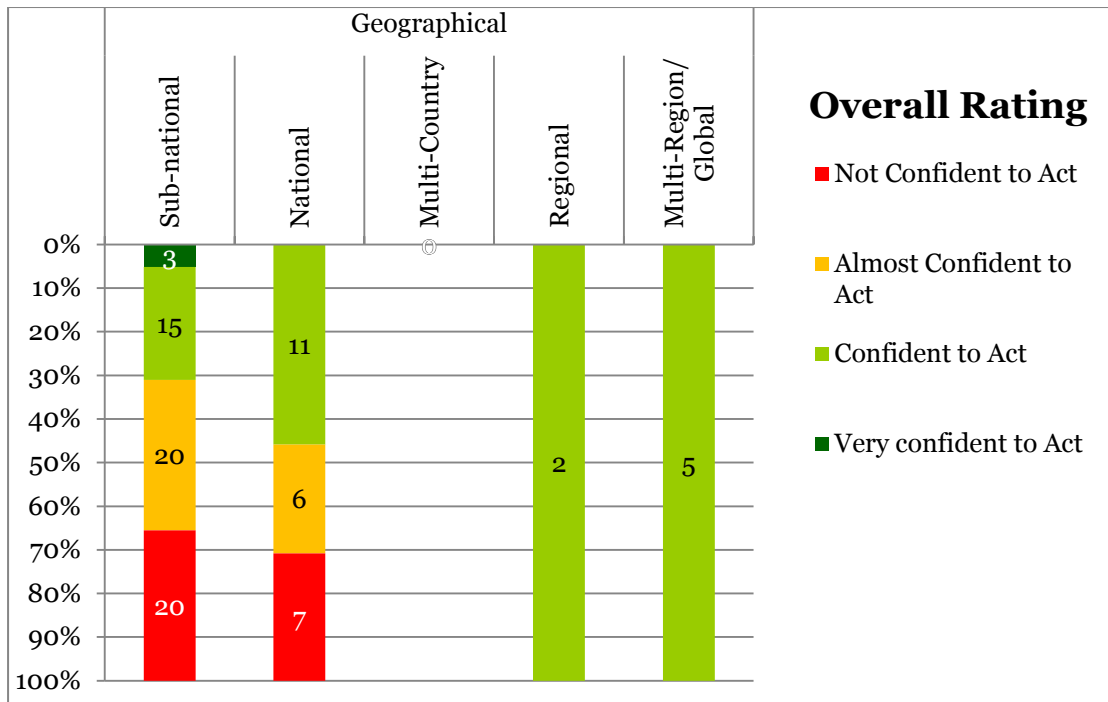




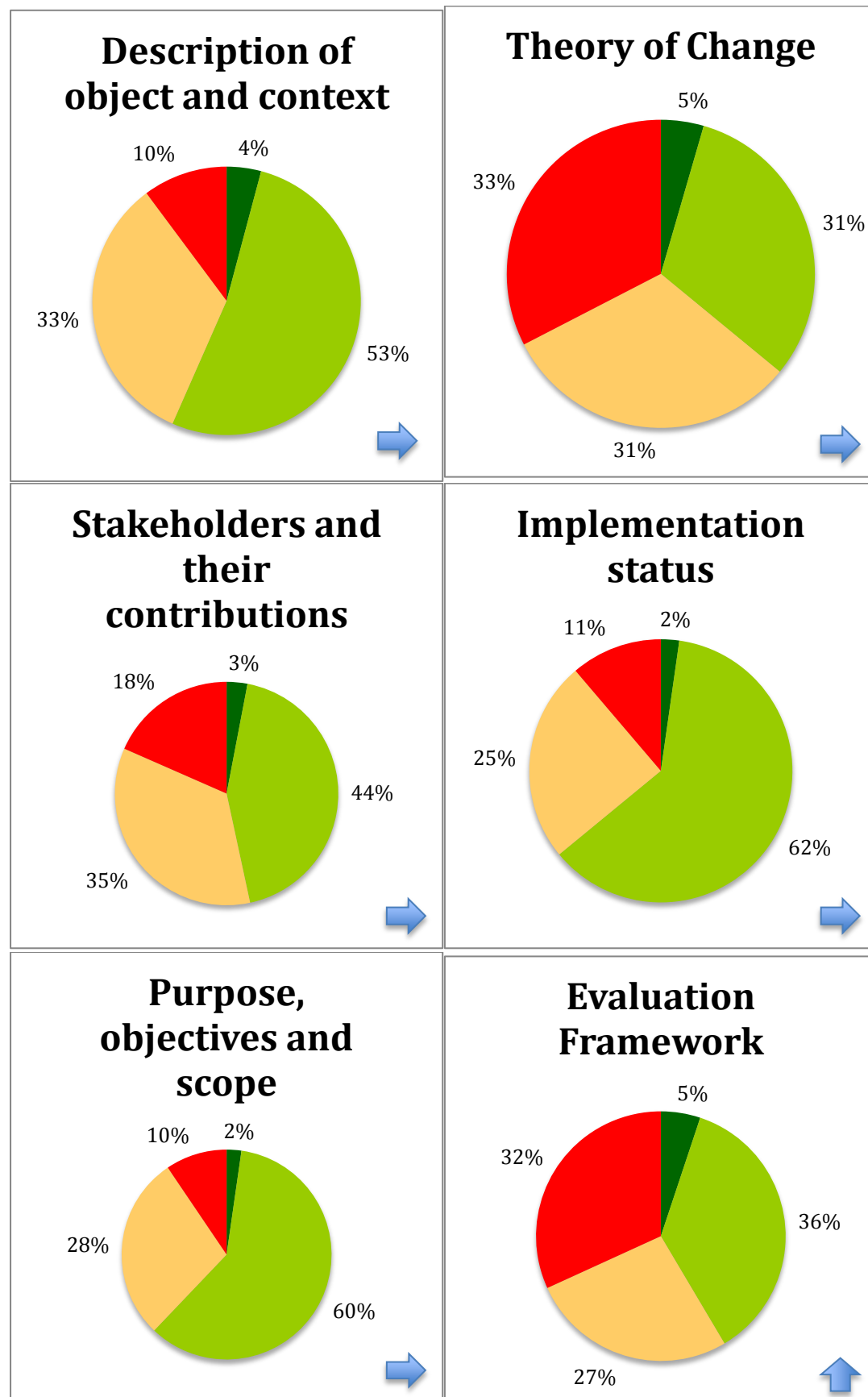
Annex 6: Ratings Graphs by Typology



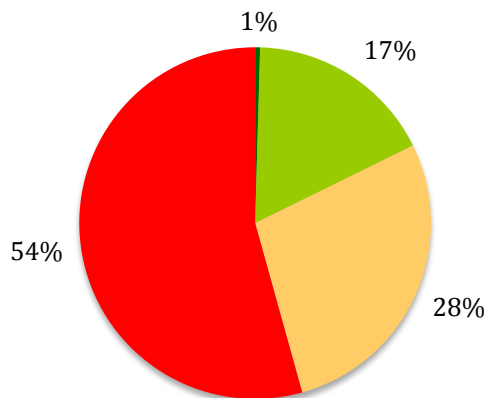




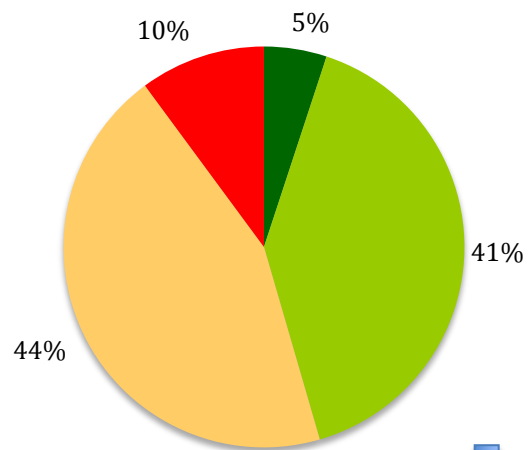
Annex 7: Performance Dashboard



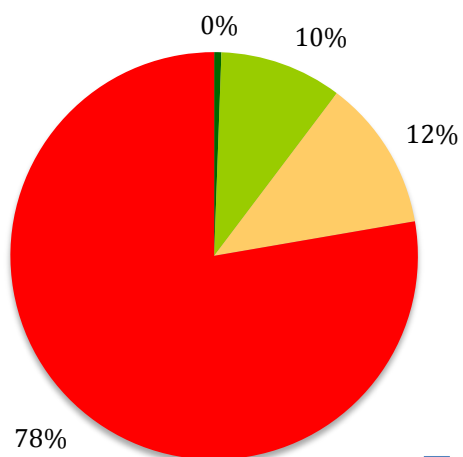
Human Rights, Gender and Equity



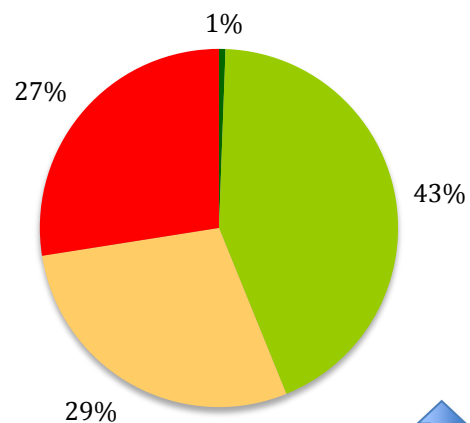
Data Collection



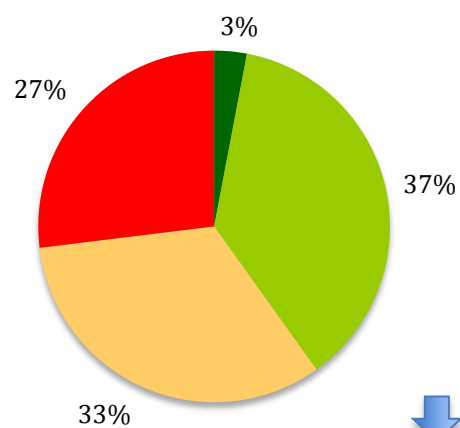
Ethics



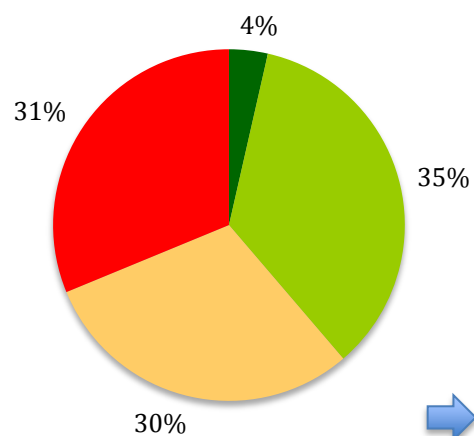
Results Based Management



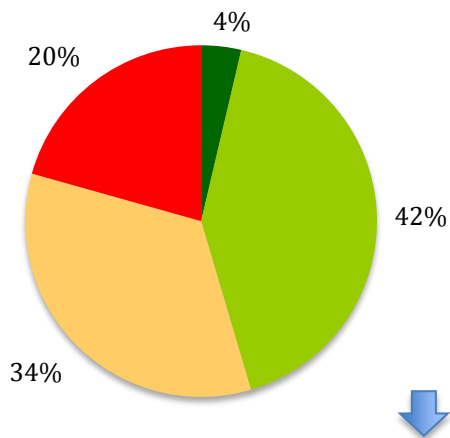
Stakeholder Participation



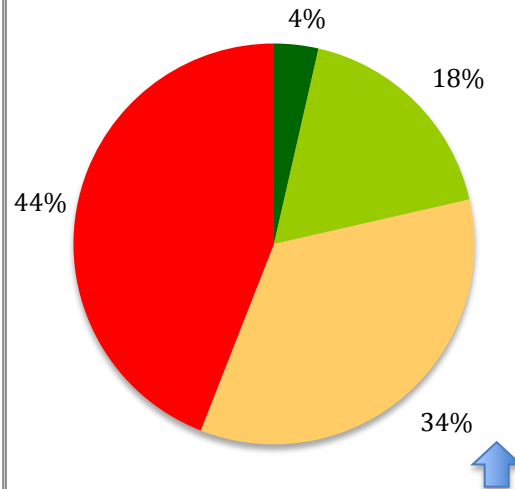
Methodological robustness



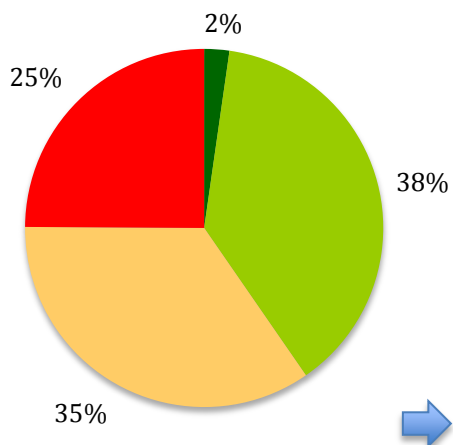
Completeness and logic of findings



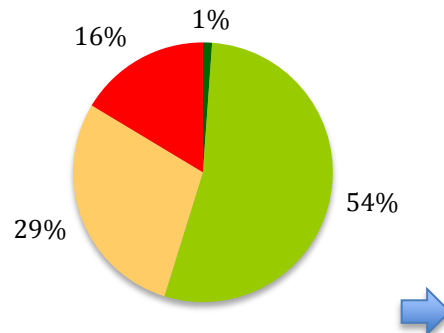
Cost analysis



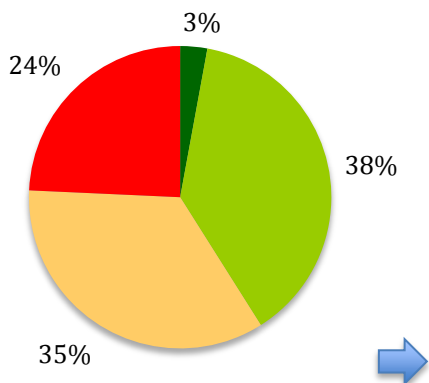
Contribution and causality



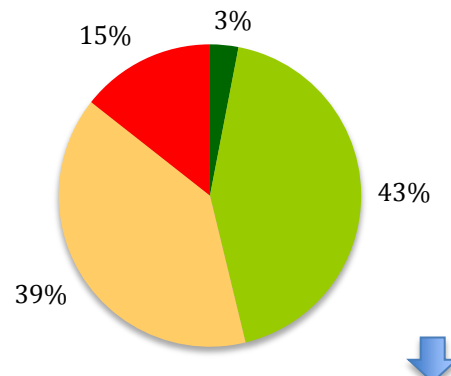
Strengths, weaknesses and future implications



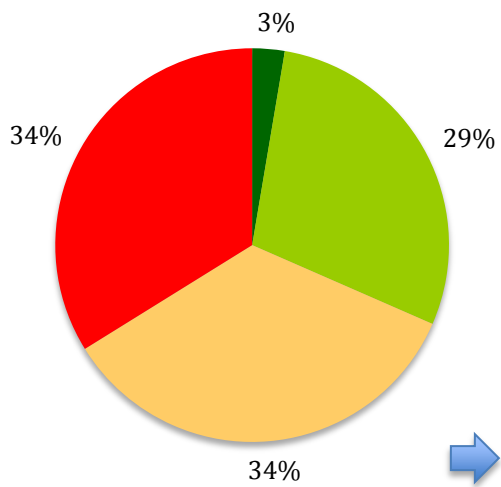
Completeness, level and insights of conclusions



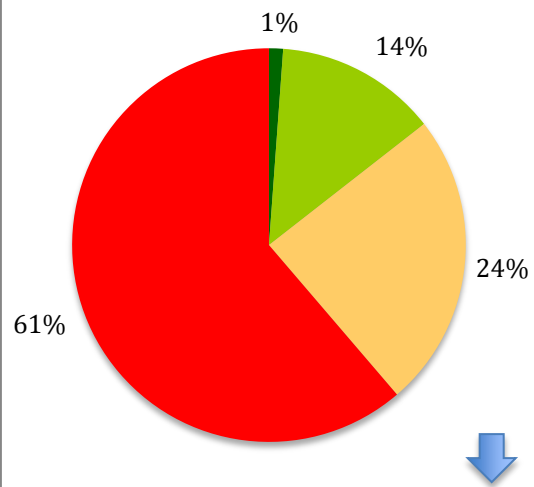
Relevance and clarity of recommendations



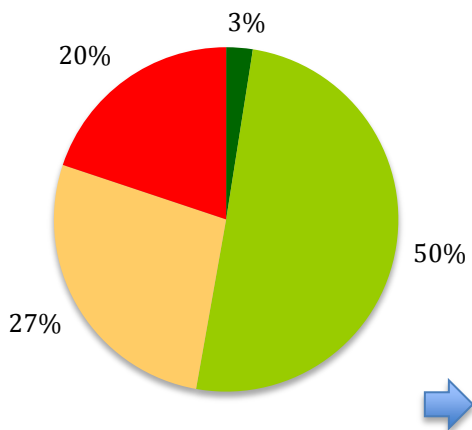
Usefulness of recommendations



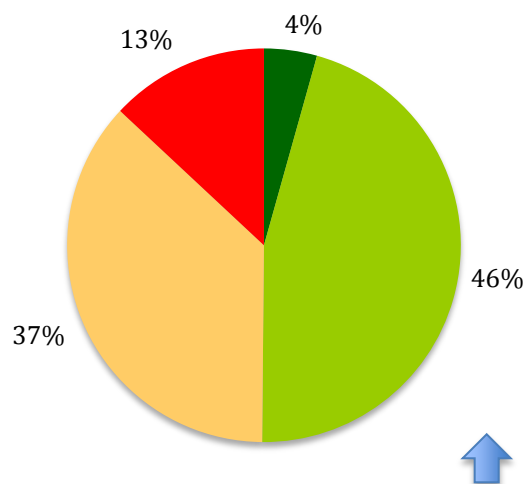
Appropriate lessons learned



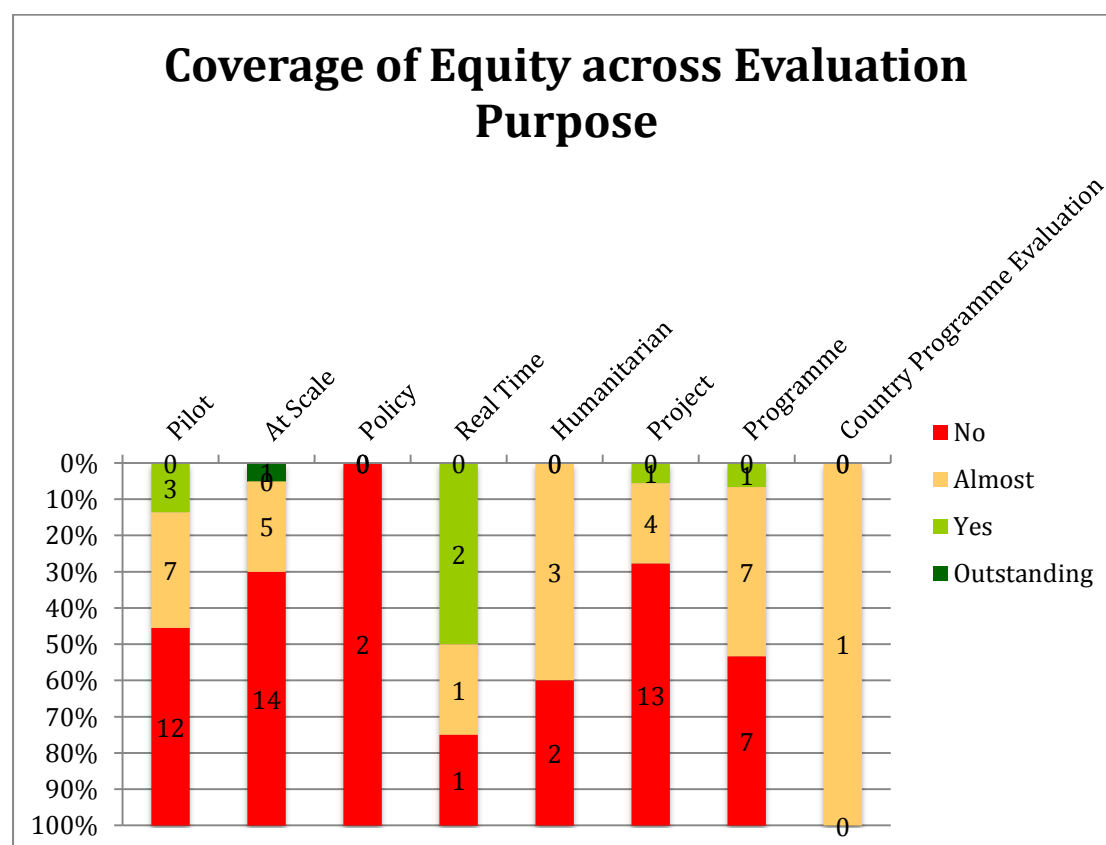
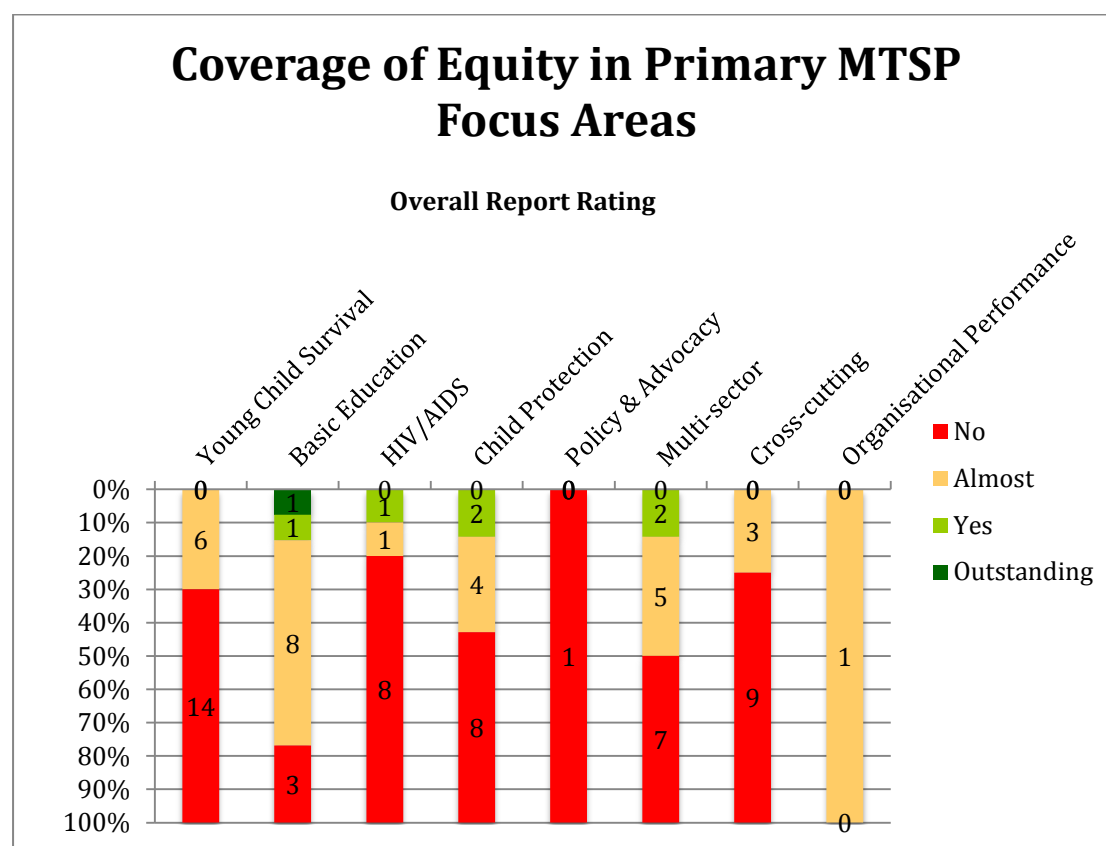
Executive Summary (Q 51, 52, 53, 54)



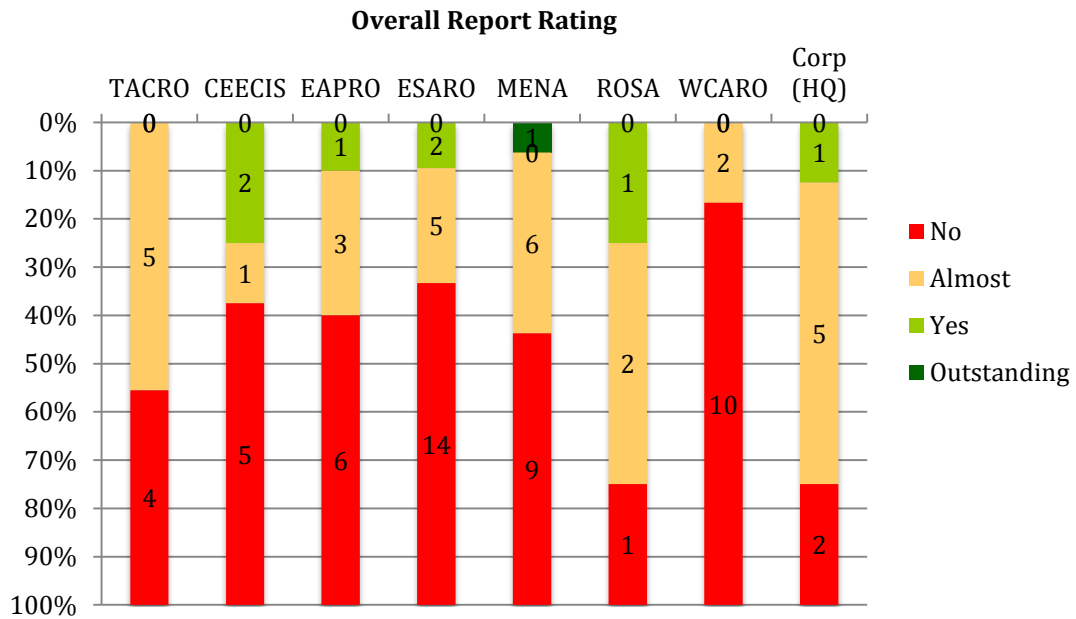
Style and presentation



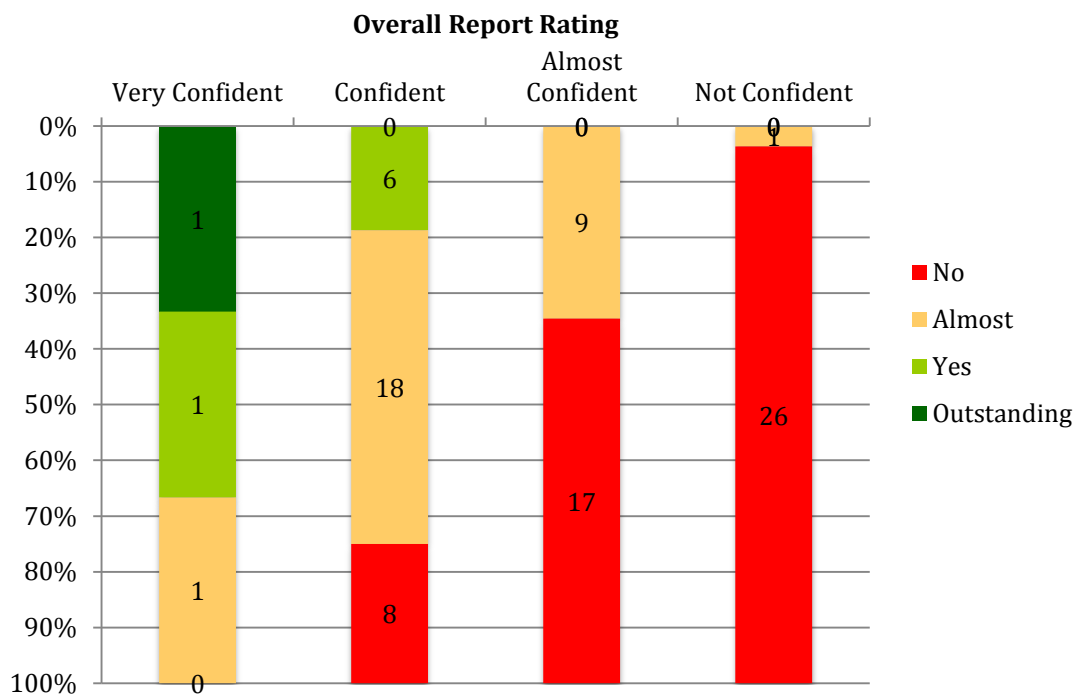
Annex 8: Additional Equity Graphs



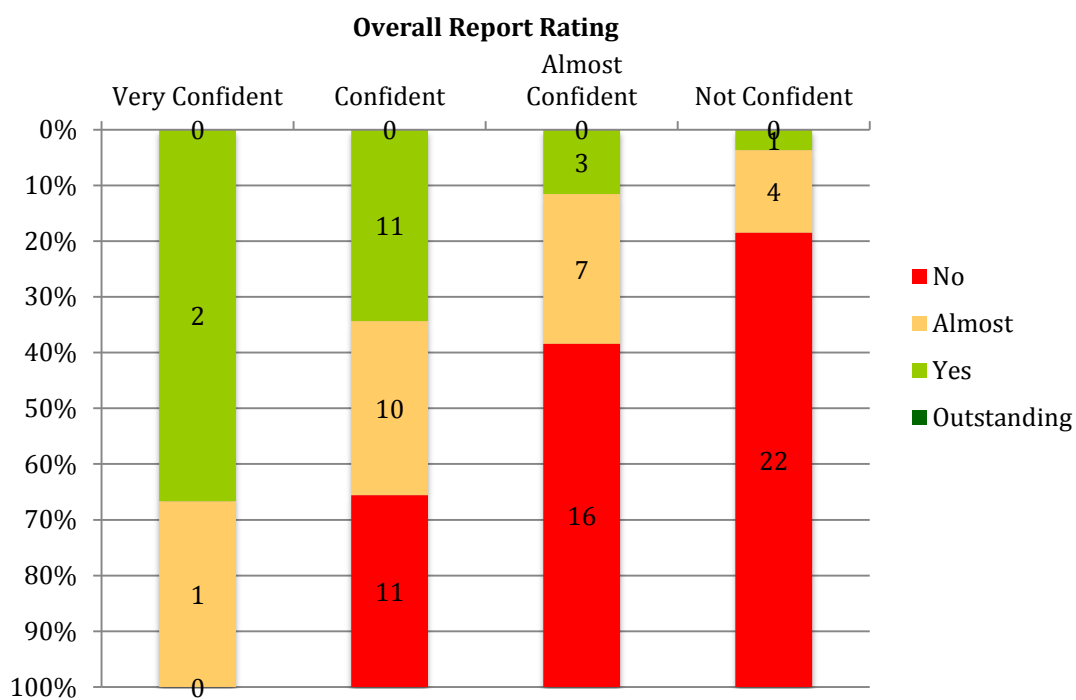
Coverage of Equity in the Regions



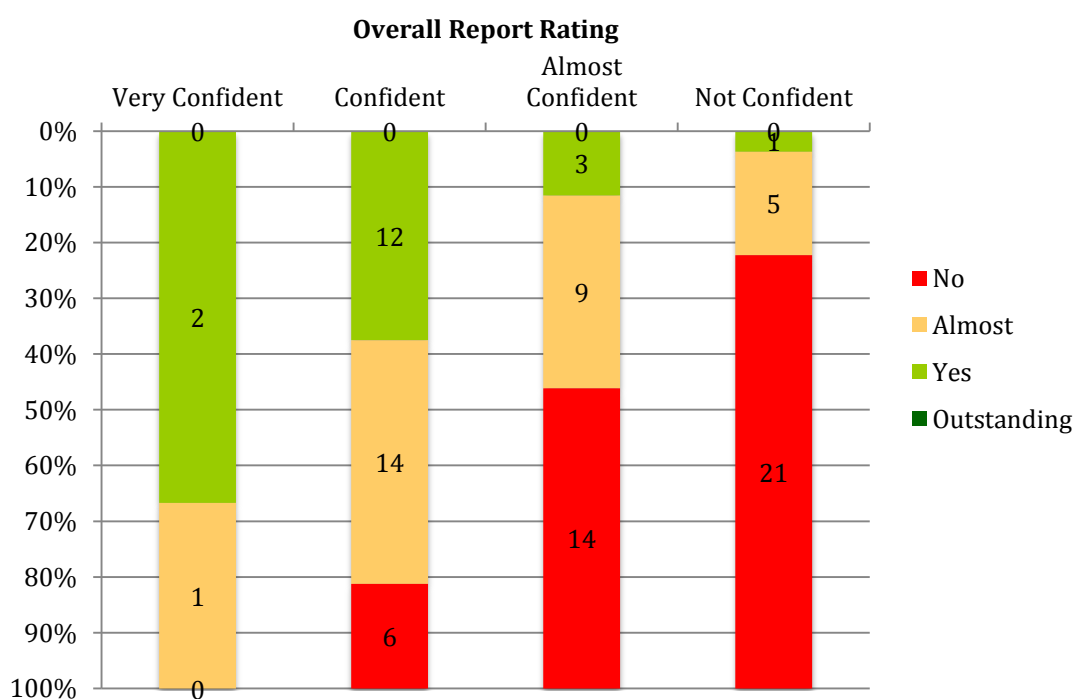
Coverage of Equity in Reports of different ratings



Coverage of Human Rights in Reports of different ratings



Coverage of GEWE in Reports of different ratings



Annex 9: Classification of reviewed 2010 reports

Number of:	Region							
	TACRO	CEECIS	EAPRO	ESARO	MENA	ROSA	WCARO	Corp. (HQ)
OVERALL								
Very confident to Act	0	0	0	1	2	0	0	0
Confident to Act	4	2	2	8	3	1	5	8
Almost Confident to Act	4	3	4	5	4	2	3	1
Not Confident to Act	1	3	4	7	7	1	4	0
SECTION A								
Very confident to Act	0	0	0	0	1	0	0	1
Confident to Act	4	4	4	9	5	3	3	7
Almost Confident to Act	4	4	4	6	8	1	9	1
Not Confident to Act	1	0	2	6	2	0	0	0
SECTION B								
Very confident to Act	0	0	0	0	1	0	1	1
Confident to Act	1	3	4	11	6	3	4	7
Almost Confident to Act	6	4	4	7	6	1	6	1
Not Confident to Act	2	1	2	3	3	0	1	0
SECTION C								
Very confident to Act	0	0	0	1	1	0	0	0
Confident to Act	4	3	2	4	5	1	4	6
Almost Confident to Act	5	2	7	10	4	3	5	2
Not Confident to Act	0	3	1	6	6	0	3	1
SECTION D								
Very confident to Act	0	0	0	1	0	0	0	0
Confident to Act	4	2	2	6	5	1	4	8
Almost Confident to Act	4	4	6	8	2	2	5	1
Not Confident to Act	1	2	2	6	9	1	3	0
SECTION E								
Very confident to Act	0	0	0	1	0	1	0	0
Confident to Act	3	1	3	6	3	0	4	8
Almost Confident to Act	5	4	6	8	7	3	2	1
Not Confident to Act	1	3	1	6	6	0	6	0
SECTION F								
Very confident to Act	0	1	0	0	1	0	0	0
Confident to Act	3	2	3	10	4	1	5	7
Almost Confident to Act	5	3	7	7	9	2	7	2
Not Confident to Act	1	2	0	4	2	1	0	0

Number of:	Geographical					Management				
	Sub-national	National	Multi-Country	Regional	Multi-Region Global	UNICEF	Joint with UN	Joint with other	Country-led	Not Clear
OVERALL										
Very confident to Act	3	0	0	0	0	2	0	1	0	0
Confident to Act	15	11	0	2	5	15	4	9	2	1
Almost Confident to Act	20	6	0	0	0	14	0	5	2	5
Not Confident to Act	20	7	0	0	0	5	1	8	1	11
SECTION A										
Very confident to Act	2	0	0	0	0	2	0	0	0	0
Confident to Act	22	13	0	0	4	18	5	11	3	1
Almost Confident to Act	26	8	0	2	1	14	0	9	0	13
Not Confident to Act	8	3	0	0	0	2	0	3	2	3
SECTION B										
Very confident to Act	2	1	0	0	0	1	0	0	0	0
Confident to Act	23	11	0	1	4	21	4	10	3	1
Almost Confident to Act	26	7	0	1	1	11	1	12	0	11
Not Confident to Act	7	5	0	0	0	3	0	1	2	5
SECTION C										
Very confident to Act	2	0	0	0	0	1	0	1	0	0
Confident to Act	15	7	0	2	5	14	3	7	2	1
Almost Confident to Act	27	11	0	0	0	17	2	9	2	7
Not Confident to Act	14	6	0	0	0	4	0	6	1	9
SECTION D										
Very confident to Act	1	0	0	0	0	0	0	1	0	0
Confident to Act	16	11	0	0	5	16	3	7	2	2
Almost Confident to Act	24	6	0	2	0	14	1	7	2	8
Not Confident to Act	17	7	0	0	0	6	1	8	1	7
SECTION E										
Very confident to Act	0	2	0	0	0	0	1	1	0	0
Confident to Act	13	9	0	2	4	12	3	8	3	1
Almost Confident to Act	26	9	0	0	1	20	1	5	2	6
Not Confident to Act	19	4	0	0	0	4	0	9	0	10
SECTION F										
Very confident to Act	1	1	0	0	0	1	0	0	1	0
Confident to Act	18	11	0	2	4	16	4	10	2	1
Almost Confident to Act	33	8	0	0	1	16	1	11	1	13
Not Confident to Act	6	4	0	0	0	3	0	2	1	3

Number of:	Purpose								Purpose 1								Purpose 2							
	Pilot	At Scale	Policy	Real Time	Humanitarian	Project	Programme	Country Programme	Pilot	At Scale	Policy	Real Time	Humanitarian	Project	Programme	Country Programme	Pilot	At Scale	Policy	Real Time	Humanitarian	Project	Programme	Country Programme
OVERALL																								
Very confident to Act	1	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Confident to Act	9	8	0	3	3	6	5	1	8	8	0	3	3	5	5	1	1	0	0	0	0	1	0	0
Almost Confident to Act	7	3	2	1	3	8	6	0	7	3	1	1	2	5	6	0	0	1	0	1	3	0	0	0
Not Confident to Act	6	8	1	0	1	10	4	0	6	8	1	0	1	8	3	0	0	0	0	0	0	2	1	0
SECTION A																								
Very confident to Act	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
Confident to Act	11	7	0	3	4	7	6	1	11	7	0	3	4	6	6	1	0	0	0	0	0	1	0	0
Almost Confident to Act	9	7	2	1	1	15	9	0	9	7	1	1	0	11	8	0	0	0	1	0	1	4	1	0
Not Confident to Act	2	5	1	0	1	2	1	0	2	5	1	0	1	1	1	0	0	0	0	0	0	1	0	0
SECTION B																								
Very confident to Act	0	0	0	1	0	0	2	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0
Confident to Act	10	11	0	3	5	9	3	1	9	11	0	3	4	7	3	1	1	0	0	0	1	2	0	0
Almost Confident to Act	9	6	2	0	2	11	11	0	9	6	1	0	2	7	10	0	0	0	1	0	0	4	1	0
Not Confident to Act	4	3	1	0	0	4	0	0	4	3	1	0	0	4	0	0	0	0	0	0	0	0	0	0
SECTION C																								
Very confident to Act	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	10	6	0	3	3	4	5	1	9	6	0	3	2	3	5	1	1	0	0	0	1	1	0	0
Almost Confident to Act	6	8	3	1	3	14	6	0	6	8	2	1	3	11	6	0	0	0	1	0	0	3	0	0
Not Confident to Act	6	5	0	0	1	6	5	0	6	5	0	0	1	4	4	0	0	0	0	0	0	2	1	0
SECTION D																								
Very confident to Act	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	9	6	0	3	3	4	8	1	8	6	0	3	3	3	8	1	1	0	0	0	0	1	0	0
Almost Confident to Act	7	9	2	1	2	12	3	0	7	9	1	1	1	9	3	0	0	0	1	0	1	3	0	0
Not Confident to Act	6	5	1	0	2	8	5	0	6	5	1	0	2	6	4	0	0	0	0	0	0	2	1	0

SECTION E																								
Very confident to Act	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	7	6	0	3	4	4	5	1	6	6	0	3	3	3	5	1	1	0	0	0	1	1	0	0
Almost Confident to Act	9	9	2	0	2	1 3	5	0	9	9	1	0	2	1 0	5	0	0	0	1	0	0	3	0	0
Not Confident to Act	7	4	1	0	1	7	6	0	7	4	1	0	1	5	5	0	0	0	0	0	0	2	1	0
SECTION F																								
Very confident to Act	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Confident to Act	9	8	0	3	3	6	6	1	8	8	0	3	3	5	6	1	1	0	0	0	0	1	0	0
Almost Confident to Act	1 1	8	2	1	4	1 3	9	0	1 1	8	1	1	3	1 0	8	0	0	0	1	0	1	3	1	0
Not Confident to Act	2	4	1	0	0	5	0	0	2	4	1	0	0	3	0	0	0	0	0	0	0	2	0	0

Number of:	Result			Level of Independence				Timing/ Stage	
	Output	Outcome	Impact	Self-Evaluation	Independent Internal	Independent External	Not Clear	Formative	Summative
OVERALL									
Very confident to Act	0	1	2	0	3	0	0	1	2
Confident to Act	5	8	18	0	14	19	0	20	12
Almost Confident to Act	8	10	8	0	9	13	4	12	14
Not Confident to Act	17	6	4	0	8	14	5	14	12
SECTION A									
Very confident to Act	1	0	1	0	1	1	0	0	2
Confident to Act	5	14	18	0	17	20	2	24	15
Almost Confident to Act	17	9	11	0	13	20	4	16	19
Not Confident to Act	7	2	2	0	3	5	3	7	4
SECTION B									
Very confident to Act	0	1	2	0	1	2	0	1	2
Confident to Act	7	12	18	0	19	19	1	22	15
Almost Confident to Act	15	9	11	0	10	20	5	19	16
Not Confident to Act	8	3	1	0	4	5	3	5	7
SECTION C									
Very confident to Act	0	0	2	0	2	0	0	0	2
Confident to Act	5	8	14	0	12	16	1	16	12
Almost Confident to Act	10	16	12	0	14	19	5	24	13
Not Confident to Act	15	1	4	0	6	11	3	7	13
SECTION D									
Very confident to Act	0	0	1	0	1	0	0	0	1
Confident to Act	4	9	17	0	13	19	0	19	12
Almost Confident to Act	13	10	9	0	12	13	7	15	17
Not Confident to Act	13	6	5	0	8	14	2	13	10
SECTION E									
Very confident to Act	0	1	1	0	0	2	0	2	0
Confident to Act	6	8	12	0	12	16	0	17	10
Almost Confident to Act	11	12	13	0	18	12	6	17	19
Not Confident to Act	13	4	6	0	4	16	3	11	11
SECTION F									
Very confident to Act	0	1	0	0	1	1	0	2	0
Confident to Act	5	11	18	0	16	19	0	21	14
Almost Confident to Act	17	12	13	0	11	24	7	19	21
Not Confident to Act	8	1	1	0	6	2	2	5	5

Number of:	MTSP Correspondence								MTSP Correspondence 1								MTSP Correspondence 2							
	Young Child	Basic	HIV/AIDS	Child	Policy &	Multi-Sector	Cross-	Organisatio	5 1	5 1	5 1	5 1	5 1	5 2	5 3	5 4	5 1	5 1	5 1	5 1	5 1	5 2	5 3	5 4
OVERALL																								
Very confident to Act	0	2	1	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	9	6	2	6	0	8	4	1	8	6	2	5	0	8	2	1	1	0	0	1	0	0	2	0
Almost Confident to Act	4	2	3	4	0	7	6	1	4	1	3	4	0	6	6	1	0	1	0	0	0	1	0	0
Not Confident to Act	9	6	4	5	1	0	5	0	8	4	4	5	1	0	4	0	1	2	0	0	0	0	1	0
SECTION A																								
Very confident to Act	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	9	6	3	8	0	9	5	1	8	6	3	7	0	9	3	1	1	0	0	1	0	0	2	0
Almost Confident to Act	8	7	4	7	0	6	6	1	8	5	4	7	0	5	6	1	0	2	0	0	0	1	0	0
Not Confident to Act	4	2	3	0	1	0	4	0	3	1	3	0	1	0	3	0	1	1	0	0	0	0	1	0
SECTION B																								
Very confident to Act	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
Confident to Act	10	6	4	7	0	7	6	1	9	5	4	7	0	7	4	1	1	1	0	0	0	0	2	0
Almost Confident to Act	8	9	3	4	0	7	6	1	8	7	3	3	0	6	6	1	0	2	0	1	0	1	0	0
Not Confident to Act	4	0	3	3	1	0	3	0	3	0	3	3	1	0	2	0	1	0	0	0	0	0	1	0
SECTION C																								
Very confident to Act	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	10	4	1	4	0	8	3	1	9	4	1	4	0	8	1	1	1	0	0	0	0	0	2	0
Almost Confident to Act	4	7	4	8	1	7	8	1	4	6	4	7	1	6	8	1	0	1	0	1	0	1	0	0
Not Confident to Act	8	4	4	3	0	0	4	0	7	2	4	3	0	0	3	0	1	2	0	0	0	0	1	0

SECTION D																								
Very confident to Act	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	7	9	1	6	0	7	4	2	6	9	1	5	0	6	2	2	1	0	0	1	0	1	2	0
Almost Confident to Act	6	2	5	6	0	8	7	0	6	0	5	6	0	8	6	0	0	2	0	0	0	0	1	0
Not Confident to Act	9	5	3	3	1	0	4	0	8	4	3	3	1	0	4	0	1	1	0	0	0	0	0	0
SECTION E																								
Very confident to Act	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Confident to Act	7	6	2	4	0	7	3	1	6	6	2	3	0	7	1	1	1	0	0	1	0	0	2	0
Almost Confident to Act	7	4	5	8	0	5	7	1	7	3	5	8	0	4	7	1	0	1	0	0	0	1	0	0
Not Confident to Act	7	6	3	3	1	2	5	0	6	4	3	3	1	2	4	0	1	2	0	0	0	0	1	0
SECTION F																								
Very confident to Act	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Confident to Act	9	7	4	6	0	8	4	1	8	6	4	5	0	8	2	1	1	1	0	1	0	0	2	0
Almost Confident to Act	1 0	7	4	8	0	7	9	1	1 0	5	4	8	0	6	8	1	0	2	0	0	0	1	1	0
Not Confident to Act	3	1	2	1	1	0	2	0	2	1	2	1	1	0	2	0	1	0	0	0	0	0	0	0