Global Evaluation Report
Oversight System

2010 Meta-evaluation

Prepared for
UNICEF Evaluation Office
1 March 2011

Joseph Barnes
Hatty Dinsmore
Sadie Watson
Purpose

• Support UNICEF decision making

• Analyse global trends and systemic issues

• Rate quality of evaluation reports

• Provide constructive feedback
Sample frame

- 96 evaluation reports reviewed from 2009
- English, French, Spanish
- All qualifying reports reviewed: 100% sample

<table>
<thead>
<tr>
<th>Region</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>TACRO</td>
<td>12</td>
</tr>
<tr>
<td>CEE/CIS</td>
<td>16</td>
</tr>
<tr>
<td>EAPRO</td>
<td>17</td>
</tr>
<tr>
<td>ESARO</td>
<td>19</td>
</tr>
<tr>
<td>MENA</td>
<td>6</td>
</tr>
<tr>
<td>ROSA</td>
<td>11</td>
</tr>
<tr>
<td>WCARO</td>
<td>12</td>
</tr>
<tr>
<td>Corporate</td>
<td>3</td>
</tr>
</tbody>
</table>
Methodology

• Based on UNICEF/UNEG Evaluation Standards
• Systematic qualitative review of evaluation report quality
• Classification by UNICEF evaluation report taxonomy
• Used standard of satisfactory report quality defined by ‘confidence to act’
Ratings

UNSATISFACTORY REPORTS
UNICEF managers *cannot* use this evaluation report with confidence

No confidence
This rated aspect falls below UNICEF evaluation report standards

Almost confident
A relatively small amount of additional work could bring this up to standard

SATISFACTORY REPORTS
UNICEF managers *can* use this evaluation report with confidence

Confident
This rated aspect meets UNICEF evaluation report standards

Outstanding
This aspect is an example of exceptional or innovative practice
Review tool

- Qualitative review of reports:
  - 58 prompting questions rated as
    - Very confident (outstanding)
    - Confident
    - Almost confident
    - Not confident
  - Comments justifying ratings
  - 6 Section ratings, comments, constructive feedback and executive feedback
  - Overall rating, comments and executive feedback
Review tool

• Classification by taxonomy
  o Geographic coverage
  o Management arrangements
  o Purpose
  o Result level
  o MTSP Correspondence
  o Level of Independence
  o Stage / Timing
Quality Assurance

- UNICEF – IOD PARC co-design workshop
- Reviewer induction, including rating test reports
- Selected peer reviewing of reviews
- Peer review of meta-evaluation report
Data processing

- Excel template used for reviews
- Excel, Word and Wordle used for analysis
- Each reviewer provided synthesis feedback
- Review dataset:
  - 1,152 pieces of report typology data
  - 6,432 individual ratings
  - 6,912 comments (140,000 words)
Limitations

• Only the report reviewed, not the evaluation
• Report-quality and not report-use assessed
• Qualitative analysis calls for judgements to be made in identifying important indicators and trends
Overall Findings

- 36% reports satisfactory
- 64% unsatisfactory
- 50% of reports could have rated satisfactory with just a little more work
Section Findings

- **OVERALL**
  - Not Confident to Act: 4
  - Almost Confident to Act: 31
  - Confident to act: 47
  - Very Confident to Act: 14

- **Object of the evaluation**
  - Not Confident to Act: 3
  - Almost Confident to Act: 49
  - Confident to act: 41
  - Very Confident to Act: 3

- **Purpose, objective and scope**
  - Not Confident to Act: 5
  - Almost Confident to Act: 45
  - Confident to act: 38
  - Very Confident to Act: 15

- **Methodology**
  - Not Confident to Act: 3
  - Almost Confident to Act: 36
  - Confident to act: 38
  - Very Confident to Act: 13

- **Findings and conclusion**
  - Not Confident to Act: 4
  - Almost Confident to Act: 36
  - Confident to act: 48
  - Very Confident to Act: 18

- **Recommendations and lessons**
  - Not Confident to Act: 3
  - Almost Confident to Act: 30
  - Confident to act: 48
  - Very Confident to Act: 14

- **Structure of the report**
  - Not Confident to Act: 4
  - Almost Confident to Act: 46
  - Confident to act: 14
  - Very Confident to Act: 14
Headlines

• 4 outstanding reports: all impact-level reports

• 6 more reports with outstanding section ratings

• Up to 3x more summative reports rated satisfactory compared to formative reports

• 90% Output-level reports unsatisfactory

• Multi-sector and cross cutting evaluations had strong reports

• Organisational performance evaluations had weak reports
Headlines

• <20% young child survival and development evaluations satisfactory

• No policy evaluations satisfactory

• Basic education and gender were strongest evaluations

• 61% reports do not include results chain

• 83% reports do not include discussion on ethics

• 86% reports included Executive Summary, but only 30% could be used with confidence
Headlines

• Only 34% attached the Terms of Reference

• Weaknesses in TORs correlate with poor report ratings

• 10% of reports praised for collecting large datasets, but most do not analyse this well

• 40 reports integrated gender considerations

• 7 reports dealt with equity

• 30% reports can HRBAP-appropriate methodologies (CFS reports strong examples)
Outstanding Reports

• Thailand’s Evaluation of Children and the 2004 Indian Ocean Tsunami

• Guinea Conakry/Guinea Bissau’s Evaluation of WASH Activities

• Timor Leste’s Evaluation of the UNICEF Education Programme

• Uzbekistan’s Evaluation of the Family Education Project
Regional ratings

<table>
<thead>
<tr>
<th>Region</th>
<th>TACRO</th>
<th>CEECIS</th>
<th>EAPRO</th>
<th>ESARO</th>
<th>MENA</th>
<th>ROSA</th>
<th>WCARO</th>
<th>Corp. (HQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td>6</td>
<td>10</td>
<td></td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall Ratings:
- Not Confident to Act
- Almost Confident to Act
- Confident to Act
- Very confident to Act
Regional ratings

**Section A Ratings**

Region

- TACRO
- CEECIS
- EAPRO
- ESARO
- MENA
- ROSA
- WCARO
- Corp. (HQ)

**Section B Ratings**

Region

- TACRO
- CEECIS
- EAPRO
- ESARO
- MENA
- ROSA
- WCARO
- Corp. (HQ)
Regional ratings

### Section C Ratings

<table>
<thead>
<tr>
<th>Region</th>
<th>TACRO</th>
<th>CEECIS</th>
<th>EAPRO</th>
<th>ESARO</th>
<th>MENA</th>
<th>ROSA</th>
<th>WCARO</th>
<th>Corp. (HQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>40%</td>
<td>50%</td>
<td>60%</td>
<td>70%</td>
</tr>
</tbody>
</table>

### Section D Ratings

<table>
<thead>
<tr>
<th>Region</th>
<th>TACRO</th>
<th>CEECIS</th>
<th>EAPRO</th>
<th>ESARO</th>
<th>MENA</th>
<th>ROSA</th>
<th>WCARO</th>
<th>Corp. (HQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>40%</td>
<td>50%</td>
<td>60%</td>
<td>70%</td>
</tr>
</tbody>
</table>
Regional ratings

**Section E Ratings**

<table>
<thead>
<tr>
<th>Region</th>
<th>TACRO</th>
<th>CEECIS</th>
<th>EAPRO</th>
<th>ESARO</th>
<th>MENA</th>
<th>ROSA</th>
<th>WCARO</th>
<th>Corp. (HQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

**Section F Ratings**

<table>
<thead>
<tr>
<th>Region</th>
<th>TACRO</th>
<th>CEECIS</th>
<th>EAPRO</th>
<th>ESARO</th>
<th>MENA</th>
<th>ROSA</th>
<th>WCARO</th>
<th>Corp. (HQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
MTSP Correspondence

Overall Ratings

- Not Confident to Act
- Almost Confident to Act
- Confident to Act
- Very confident to Act
Conclusions

• Evaluation reports benefit from having access to relevant and well-developed international frameworks
  o E.g. Child Friendly Schools
  o This helps evaluators to integrate rights and results

• A disjuncture exists between successful evaluation and strong integration of rights
  o It seems that the right blend of skills is not available, so rights evaluations rely on technical specialists
  o This is in evidence particularly in relation to HIV/AIDS evaluations
Conclusions

• The evaluation function is not delivering consistent contributions to upstream knowledge management
  o Policy evaluation is a notably weak area, particularly in recommendations and lessons learned
• Robust and transparent analysis of data is a problem
  o Evaluation reports are stronger in the theory stages of the review, but then struggle to maintain logical progression from data through analysis to findings and conclusions
Conclusions

• Weak terms of reference are contributing to poor report quality
  ○ Reports tend to build of TORs and also reflect the weakness within them, particularly around unclear purposes, objectives and evaluation frameworks

• Fundamental misunderstandings of lessons learned prevail
  ○ Lessons learned tend to be contextual to the evaluated object or concern operational changes that need to be made
Conclusions

• The qualitative approach is a viable and useful way forward for the UNICEF Global Evaluation Report Oversight System

• There is work to be done in supporting evaluation managers to mainstream human rights, gender and equity
  - Challenges remain to integrating strong HRBAP and strong evaluation consistently
  - The meta-evaluation tool can also be refined to investigate this further
Lessons Learned

• There is great value to blending the skills of evaluation teams
  o Evaluation skills and sector knowledge can be brought together just as international and national knowledge currently is

• Developing strong international frameworks provides a platform for stronger, more rights orientated and more useful evaluation

• Co-designing the methodology and investing in the development stage of evaluation-quality-reviews delivers a strong return in performance
Recommendations

• UNICEF Evaluation Office
  o Focus on delivering a strategy for more consistently high quality terms of reference within the decentralised evaluation function
  o Continue development of the qualitative approach to GEROS
  o Develop and communicate an integrated HRBAP mainstreaming strategy for the evaluation function
  o Revisit how requirements for recommendations and lessons learned sections are communicated to evaluators and managers
Recommendations

• Regional and Country Offices together
  o Cooperate on delivering basic quality assurance at the TOR and draft report stages
  o Attempt to focus on fewer and better evaluations that deliver strategic priorities
  o Create cross-UNICEF pollination around upstream evaluations, and explore options for multi-country approaches
  o Invest in clarifying results frameworks for challenging thematic areas, work to evaluate these at higher levels, and contribute this knowledge to international attempts to create coherent frameworks
  o Deliver extra support to country-led evaluations
# Regional areas of focus

<table>
<thead>
<tr>
<th>Region</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>TACRO</td>
<td>Focus on enhancing the processes for developing and communicating recommendations that are used by evaluators</td>
</tr>
<tr>
<td>CEE/CIS</td>
<td>Focus on maximising experience with rights issues by increasing the involvement of evaluators and tackling higher-level results</td>
</tr>
<tr>
<td>EAPRO</td>
<td>Learn from the CFS multi-country evaluations to more consistently structure evaluation reports and eradicate unsatisfactory conclusions sections</td>
</tr>
<tr>
<td>ESARO</td>
<td>Focus on enhancing the essential ingredients of a good evaluation through improving the quality of terms of reference</td>
</tr>
<tr>
<td>MENA</td>
<td>Invest in developing access to a wider cohort of evaluators with relevant contextual knowledge and experience</td>
</tr>
<tr>
<td>ROSA</td>
<td>Focusing on quality assuring the analytical capability of evaluators by requiring detailed methodologies to be provided in inception reports (including analytical frameworks and processes)</td>
</tr>
<tr>
<td>WCARO</td>
<td>Explore options for sharing and learning from regional examples that are outstanding, particularly in relation to mixed methods</td>
</tr>
<tr>
<td>Corporate</td>
<td>Focus on sharing lessons for consistent evaluation with regions and try to maintain this consistency with a higher rate of knowledge generation</td>
</tr>
</tbody>
</table>
Online resources


UNICEF Evaluation Policy [English]

UNICEF/UNEG Evaluation Report Standards [INTRANET]

Presentation on UNICEF Global Evaluation Report Oversight System [INTRANET]

UNICEF/UNEG Terms of Reference Standards [INTRANET]

Technical Note 1: Children Participating in Research and Monitoring and Evaluation (M&E) - Ethics and Your Responsibilities as a Manager

Technical Note 3: Writing a good Executive Summary