



United Nations  
Educational, Scientific and  
Cultural Organization

Internal Oversight Office  
Evaluation Section

IOS/EVS/PI 168

# 2017 SYNTHESIS REPORT

January 2018

## Abstract

This report documents the comparative analysis and quality assessment of 5 corporate and 12 decentralized evaluations. All evaluations stem from 2017. The report sheds light on UNESCO's performance, both content-wise and in evaluation.

The report concludes that UNESCO's work is relevant, but that its performance can be improved. To that end, the IOS Evaluation Office suggests greater focus and selectivity in UNESCO's programming, better alignment with national-level capacities and enhanced qualitative inquiry into UNESCO's results.

UNESCO's evaluation reports, whether on corporate or decentralized evaluations, are quite good, but tend to be too descriptive and too long. They need to focus more on the usage of evaluation outcomes by the evaluation's primary users. They also need to become more analytical to foster understanding and learning.

Based on these findings, the IOS Evaluation Office will (i) strengthen its own evaluation and UNESCO internal advisory practice; and (ii) advise UNESCO's senior management, the Oversight Advisory Committee, the program sectors, field offices and the category 1 institutes on how to further improve UNESCO's performance.

## Acronyms

IDP	Internally Displaced Persons
IOS	Internal Oversight Service
OAC	Oversight Advisory Committee of IOS
OECD	Organization for Economic Cooperation and Development - Development Assistance Committee
RBM	Results-based Management
UNEG	United Nations Evaluation Group
UNESCO	United Nations Educational, Scientific and Cultural Organization

## Table of Content

Executive summary .....	3
<b>1.</b> Introduction .....	5
<b>2.</b> Basic portfolio data.....	7
<b>3.</b> Quality of work.....	9
<b>4.</b> Quality of evaluations.....	13
Appendix: Quality of the evaluation reports – a scorecard.....	19
References .....	26

## Tables

Table 1. Scoring of evaluands on the standard evaluation criteria.....	10
Table 2. Quality assessment of UNESCO’s evaluation reports.....	14
Table 3. Basic information requirements of evaluation reports.....	15
Table 4. Detailed scoring of the evaluation reports. ....	19

## Figures

Figure 1. Thematic and geographical spread of the evaluations. ....	7
Figure 2. Sector coverage of the evaluations .....	8
Figure 3. Type of evaluation .....	8
Figure 4. Affiliation of evaluators.....	8
Figure 5. Report length .....	8
Figure 6. Evaluation reports' coverage of the standard evaluation criteria .....	9
Figure 7. The 7 dimensions of a quality evaluation report.....	13

## Textbox

Textbox 1. The applicability of the standard evaluation criteria .....	9
Textbox 2. The applicability of the UNEG Quality Checklist of Evaluation Reports	13

## Executive summary

This synthesis report answers two questions:

1. How does UNESCO perform against the evaluation criteria: relevance, effectiveness, efficiency and sustainability (and impact)?
2. What is the quality of UNESCO's evaluation reports (when measured against the UNEG Quality Checklist for Evaluation Reports)?

The answers to these two questions reveal where UNESCO performs well and where there is room for improvement. Based on these insights, the IOS Evaluation Office:

1. informs UNESCO's senior management and Oversight Advisory Committee on the strength and challenges of UNESCO;
2. advises the program sectors, field offices and category 1 institutes on how to improve their performance;
3. enhances its own corporate evaluation practice; and
4. sharpens its advice and support on decentralized evaluations.

This synthesis report covers 5 corporate and 12 decentralized evaluations from 2017. The report's findings, conclusions and recommendations rest on an analysis of the evaluation reports.

**How does UNESCO perform against the evaluation criteria: relevance, effectiveness, efficiency and sustainability (and impact)?**

All evaluations judge UNESCO's work to be *relevant*. UNESCO addresses clear and undisputed development needs which are acknowledged and targeted by the respective governments. At the global level, UNESCO is appreciated for providing a neutral platform for national action and international cooperation.

None of the evaluation reports make definite statements on the evaluands' *effectiveness*, i.e. on the attainment of medium- and long-term outcomes. The weak application of (i) monitoring and evaluation principles by UNESCO's programs and projects; and (ii) Theory of Change and contribution analysis by the evaluators hamper conclusive statements on this criterion. Generally, the more thematically- focused interventions appear most promising.

In judging *efficiency*, the reports reflect on organizational and implementation issues. Most evaluation reports point to the professionalism of UNESCO staff and implementation partners and the prudent use of resources. The corporate evaluations warn about understaffing, spreading resources too thinly, over-institutionalization of global initiatives, and underutilization of partnerships.

All evaluation reports point towards the continued need for stewardship to secure the *sustainability* of results. They stress the importance of continued buy-in and ownership by national actors. Many UNESCO supported interventions struggle to compete for national funding with '*more acute*' development challenges.

Looking forward, the IOS Evaluation Office recommends the program sectors, field offices and category 1 institutes to:

1. focus their scarce financial and human resources on a limited set of interventions;
2. select intervention areas where national-level stakeholders can lead, both in the short and long term;
3. assess the effectiveness of UNESCO's work and foster learning through the enhanced use of Theories of Change and evaluative analysis.

**What is the quality of UNESCO's evaluation reports (when measured against the UNEG Quality Checklist for Evaluation Reports)?**

Generally, the evaluation reports score better on the *framing* of the evaluation and evaluand than on the *evaluative analyses*. There is room for improvement on all quality dimensions.

The *purpose statements* of the evaluations tend to be generic and fail to identify the intended users of the evaluation, why they needed the evaluation, what their information needs were, and how the evaluation results will be used. The user orientation of UNESCO's evaluations can be significantly strengthened.

The reports are all *logically structured* and (more or less) cover the basic data requirements of the UNEG Quality Checklist. The checklist offers a more structured way of presenting these basic data requirements.

All evaluations provide a comprehensive *description of the evaluand*. The challenge is brevity. The IOS Evaluation Office will promote a 2-page maximum to this end.

On *methodology*, most evaluations score well on the description of the data collection methods and data sources. Half of the evaluations include the limitations faced by the evaluation. Few evaluations explain the design of the evaluation and how the data sets and methods provide for a representative picture of the evaluand.

Five evaluations score well on their *evaluative analysis*. Most evaluation reports struggle however on this front. They are too descriptive and insufficiently argue their findings and conclusions through a purposeful, structured and coherent presentation and analysis of key evidence against pre-defined benchmarks.

All evaluation reports include *recommendations* which are relevant to the evaluand. The recommendations are mostly provided in a separate chapter which weakens the linkage with their evidence-base (exacerbated by the sheer length of many reports). Only six evaluations include the target group for (individual) recommendations.

To further improve UNESCO's evaluation reports, these reports need to:

1. Focus more on the use of the evaluation by the primary intended users by soliciting and explicitly answering their key evaluation questions and more purposeful evaluation designs;
2. Become more analytical and learning-oriented by clearly distinguishing between data and findings, comparing findings with pre-defined benchmarks, and reflecting on the outcome of this comparative analysis;
3. Strengthen the gender perspective in the evaluation design phase, especially for decentralized evaluations and argue when the adoption of a gender perspective is unnecessary.

Based on the above assessment, the IOS Evaluation Office has prepared a new guideline and template for UNESCO's corporate and decentralized evaluation reports. This new guideline is available in a separate document.

## 1. Introduction

1. This report documents the comparative analysis and quality assessment of 5 corporate and 12 decentralized evaluations of 2017.<sup>1</sup> This opening chapter lays down the purpose, use, evaluation questions, approach and limitations of the review. It also explains the structure and scope of the report.

### 1.1 Purpose and use

2. This report sheds light on UNESCO's performance, both content-wise and in its evaluation practice. The report reveals where UNESCO performs well and where there is room for improvement. Based on these insights, the IOS Evaluation Office:

1. informs UNESCO's senior management and Oversight Advisory Committee on the strength and weaknesses of UNESCO's work and evaluation practice;
2. advises the program sectors, field offices and category 1 institutes on how to improve their performance;
3. enhances its own corporate evaluation practice; and
4. sharpens its advice and support on decentralized evaluations.

3. The IOS Evaluation Office is the *primary user* of this report. The IOS Evaluation Office will nonetheless pro-actively share the review – during its regular consultative meetings – with the Oversight Advisory Committee, program sectors, field offices and category 1 institutes. The latter groups constitute the *secondary users* of this report.

### 1.2 Evaluation questions

4. This review answers two questions:

1. How does UNESCO perform against the evaluation criteria: relevance, effectiveness, efficiency and sustainability (and impact)?

2. What is the quality of the evaluations when measured against the UNEG Quality Checklist for Evaluation Reports (UNEG 2010)?

### 1.3 Approach

5. The IOS Evaluation Office conducted a *desk-review* of 5 corporate and 12 decentralized evaluations. The review constituted a comparative analysis and quality assessment of the reports. The IOS Evaluation Office rated UNESCO's performance and the evaluation reports' compliance with the UNEG quality standards with a *streetlight model*: a **green light** when the evaluation criteria or quality standard are *fully met*; an **orange light** when these are *partially met*; and a **red light** when they are *not met*. More importantly, the IOS Evaluation Office identified – through an inductive analysis – *common themes* which influence the attainment of the evaluation criteria or quality standards.

6. This report includes:

1. for each evaluation, the evaluation of UNESCO's performance against the evaluation criteria and report quality standards;
2. the emerging themes from the comparative analysis of the 17 reports;
3. recommendations on how UNESCO can improve its performance, both content-wise and in its evaluation practice.

### 1.4 Limitation

7. The report's findings, conclusions and recommendations rest on a *document analysis*, i.e. the structured review of the evaluation reports. We did not interview the external evaluators or the evaluand's program officers.

---

<sup>1</sup> The 12 decentralized evaluations are those received by the IOS Evaluation Office on 1 December 2017 (the cut-of date for this review). Since, 10 more decentralized evaluations have reached the IOS Evaluation Office.

## 1.5 Reading guide

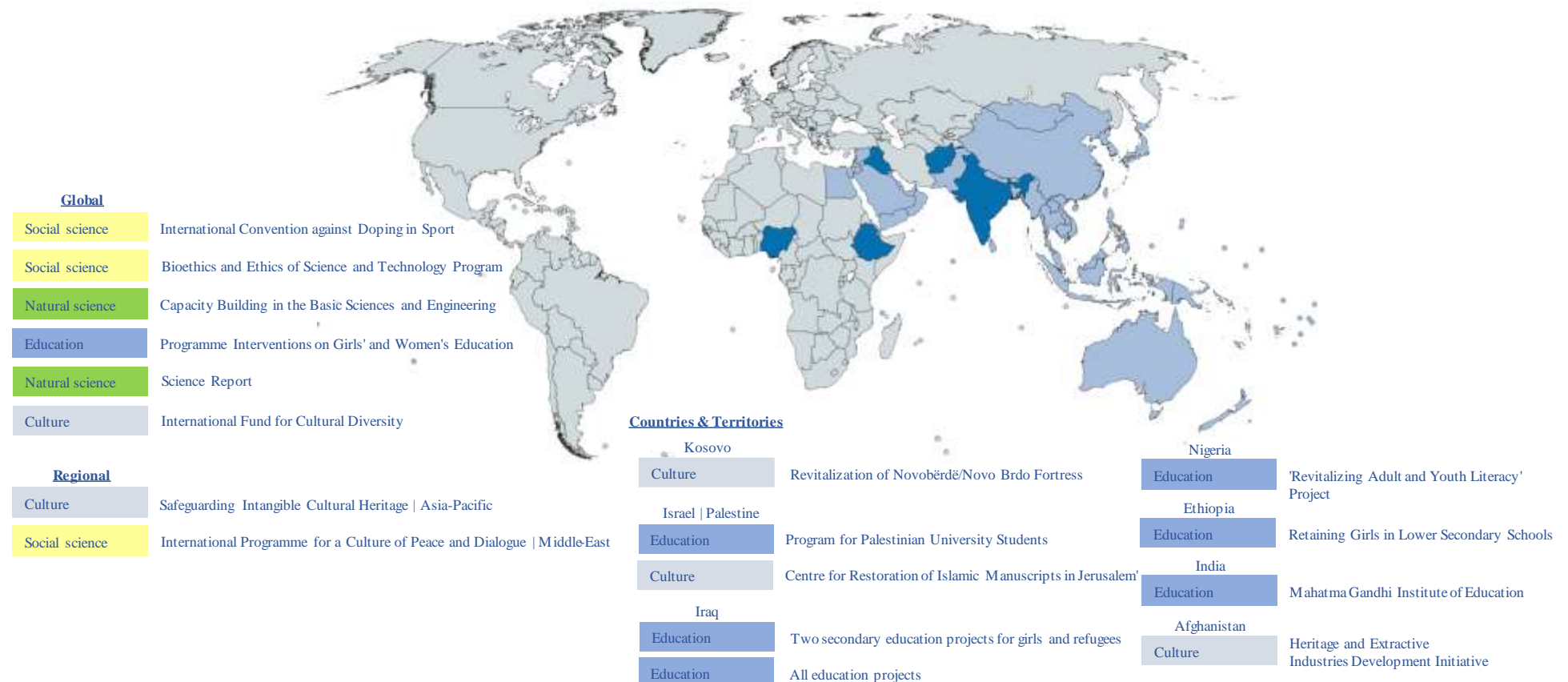
8. The next chapter presents basic data on the 5 corporate and 12 decentralized evaluations. Chapter 3 and 4 answer the two core evaluation questions respectively and provide recommendations on how UNESCO can improve its performance.

## 2. Basic portfolio data

9. This Chapter briefly introduces the 5 corporate and 12 decentralized evaluations. Figure 1 lists all evaluations, including their thematic focus, and shows their geographic coverage. The evaluations are split evenly between evaluands with a global/regional and a country focus. Figure 2 on the next page shows that the education and culture

sectors provided most of the evaluations. The evaluations are equally split between mid-term and ex-post evaluations (Figure 3). All evaluations were conducted by external evaluators – roughly split evenly between individual consultants and evaluation firms (see Figure 4). Finally, Figure 5 shows the average and variation in the length of the evaluation reports.

Figure 1. Thematic and geographical spread of the evaluations.



Map source: <https://mapchart.net/detworld.html>



Figure 2. Sector coverage of the evaluations

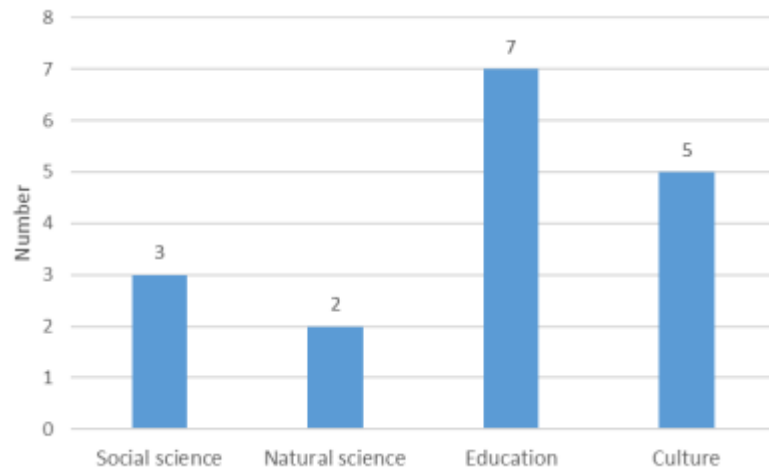


Figure 3. Type of evaluation

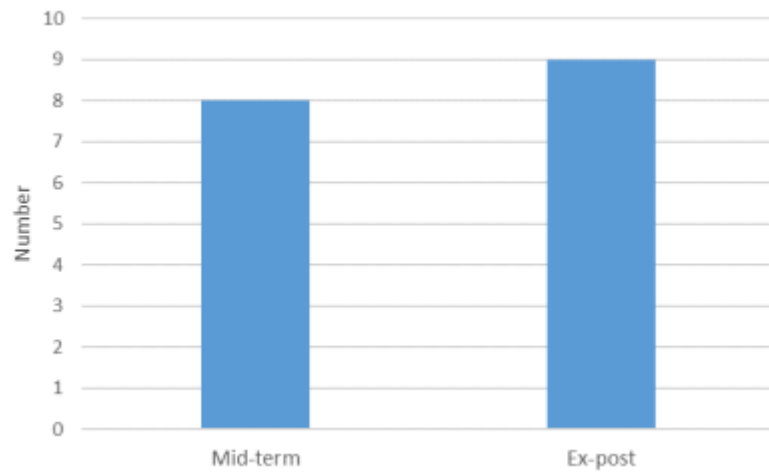


Figure 4. Affiliation of evaluators

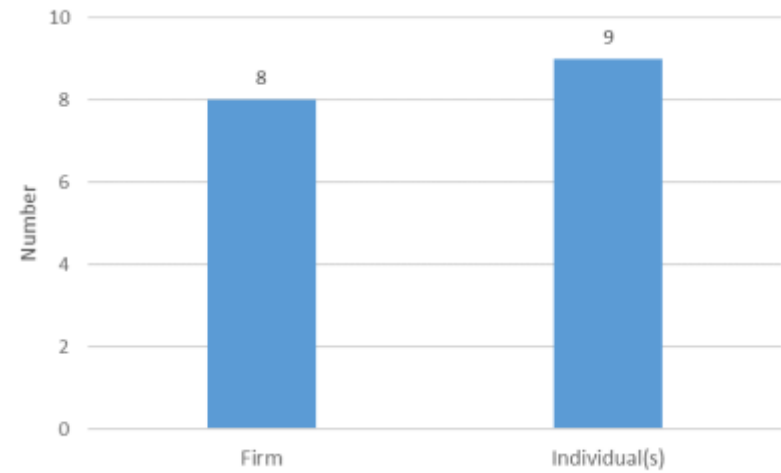
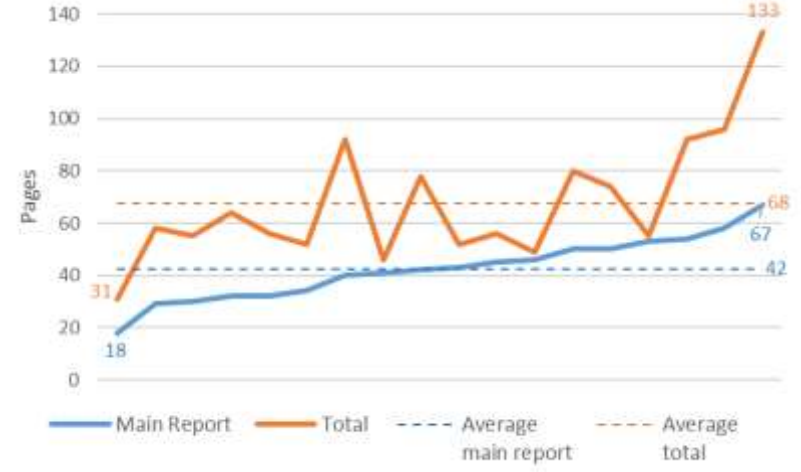


Figure 5. Report length



### 3. Quality of work

*How does UNESCO perform against the evaluation criteria: relevance, effectiveness, efficiency and sustainability?*

#### 3.1 Introduction

10. This chapter's question implicitly assumes that the underlying evaluation reports cover the standard evaluation criteria. Coverage of these criteria is not an ex-ante requirement (see Textbox 1). Figure 6 shows that most reports address some or all evaluation criteria. Few reports however include explicit ratings or definitive statements on the extent to which the evaluands meet the criteria. In most cases, we inferred the ratings from the reports' analysis.

11. Table 1 on the next page shows how each evaluation report scores on the standard evaluation criteria (using the streetlight model explained in Chapter 1). Two observations come to the fore. First, UNESCO is judged to be *relevant* by all evaluations. Second, no evaluand receives a green light on all evaluation criteria and just a handful obtain three green lights (out of a maximum of four). In the subsequent sections, we reflect on the scoring per evaluation criterion. We then draw an overall conclusion and formulate recommendations for the program sectors, field offices and the category 1 institutes.

#### 3.2 Relevance

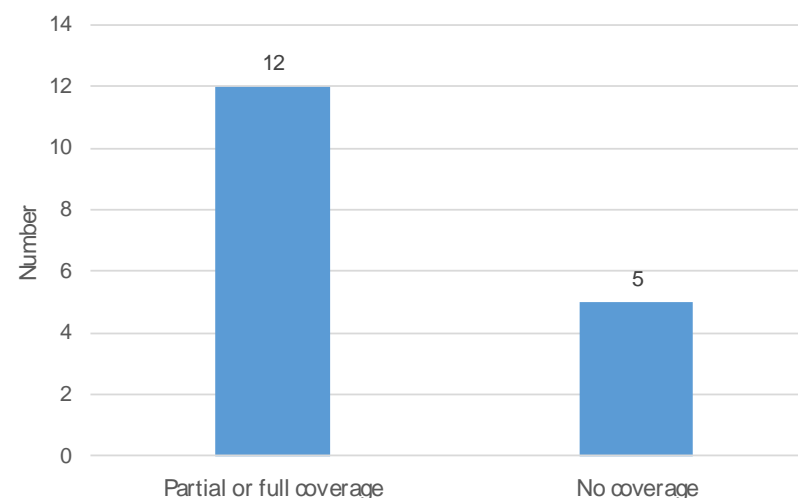
12. Relevance concerns *'the extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor'* (OECD n.d.). UNESCO scores well on this evaluation criterion. At the country level, UNESCO addresses clear and undisputed development needs, such as – amongst others – literacy training, access to educational material, harnessing (intangible) cultural heritage or access to scientific knowledge and trends.

13. These development needs are explicitly acknowledged and targeted by the respective governments even when – often – it allocates little human and financial resources to their address. At times, UNESCO fills a gap left by government or other

#### Textbox 1. The applicability of the standard evaluation criteria

The evaluation criteria *relevance, effectiveness, efficiency and sustainability* are commonly used in program evaluations and have been codified by the OECD-DAC (OECD n.d.). UNESCO recognizes these criteria and has included them in its Evaluation Policy 2014 – 2021 (UNESCO 2015). They form however *guiding principles* and need not be included in every evaluation. The ultimate selection of an evaluation's evaluation criteria is a function of: (i) the information requirements of the intended users of the evaluation; (ii) the available resources for the evaluation (in terms of money, staff and time); and (iii) the evaluability of the evaluand along these criteria.

Figure 6. Evaluation reports' coverage of the standard evaluation criteria



actors. Whilst good in the short-run, this begs the question as to the sustainability of the efforts – a topic we return to in section 3.5.

14. The intervention areas are aligned with the Sustainable Development Goals (SDGs), UNESCO supported international conventions, and UNESCO biannual Programme and Budget documents.

Table 1. Scoring of evaluands on the standard evaluation criteria

		Standard evaluation criteria included?	Relevance	Effectiveness	Efficiency	Sustainability
#	Corporate evaluations					
1	International Convention against Doping in Sport	No			n/a	n/a
2	Bioethics and Ethics of Science and Technology Programme	No			n/a	n/a
3	Work in Capacity Building in the Basic Sciences and Engineering	Yes				
4	Programme Interventions on Girls' and Women's Education	Yes				
5	Science Report	Yes				
	Decentralized evaluations					
1	'Revitalizing Adult and Youth Literacy' Project	No				n/a
2	Program for Palestinian University Students under Conditions of Severe Poverty	Yes				
3	Revitalization of Novobërdë/Novo Brdo Fortress Project	Yes				
4	Mahatma Gandhi Institute of Education for Peace and Sustainable Development	No				
5	International Fund for Cultural Diversity	Yes			n/a	
6	Two education projects: 'Expanding secondary education for Syrian refugees' and 'Bridging the gaps in secondary schooling for girls of IDP communities'	Yes				
7	'Afghanistan Heritage and Extractive Industries Development Initiative'	No		n/a	n/a	n/a
8	Project 'Safeguarding Intangible Cultural Heritage through the Strengthening of National Capacities in Asia and the Pacific'	Yes				
9	Project 'Retaining Girls in Lower Secondary Schools and increasing their Learning Outcomes in Gambella and Somali Regional States'	Yes				
10	Education Projects in Iraq	Yes				
11	International Programme for a Culture of Peace and Dialogue	Yes				n/a
12	Project 'Ensuring the sustainability of the Centre for Restoration of Islamic Manuscripts in Jerusalem'***	Yes				

Legend: n/a = not assessed by evaluation

\* This concerns a self-evaluation by the UNESCO field office with two separate external validations on the technical and institutional aspects of the UNESCO project respectively. We have assessed the two separate external validation reports as one. The validation report on the technical aspects of the UNESCO project includes an assessment of the standard evaluation criteria, whereas the validation report on the institutional aspects does not include such an assessment.

15. Almost half of the evaluations concern global initiatives, frequently embedded in international conventions. Here, UNESCO is considered particularly relevant. UNESCO is appreciated for its neutrality, global reach, as well as for providing 'an intergovernmental platform' or 'a critical point of reference' for national action and international exchange and cooperation.

### 3.3 Effectiveness

16. Effectiveness is 'a measure of the extent to which an aid activity attains its objectives' (OECD n.d.). Effectiveness concerns attainment of medium- or long-term outcomes. This constitutes a challenging criterion for UNESCO. As one evaluation report states:

*'Measuring, monitoring and evaluating outcomes and impacts beyond the output level remains challenging ... due to a lack of adequate monitoring and evaluation frameworks.'*

17. None of the evaluands are undergirded by a Theory of Change.<sup>2</sup> Although most evaluations have some form of Results or Logical Framework, these do not spell out how an intervention will achieve its medium- to long-term outcomes.<sup>3</sup> Moreover, most evaluands do not provide for systematic data collection and monitoring. None of the evaluation reports make definite statements on the achievements of objectives at the medium- and long-term outcome level.

18. The question thus remains to what extent success at the output level – providing access to libraries, renovating a historically unique fortress, or empowering State Parties to address doping – will affect sustainable development, such as: improved education performance, sustainable tourism or reduced use of doping in sport. The uneven application of Results-Based Management principles is a systemic issue within UNESCO. It requires continued attention.

---

<sup>2</sup> Some evaluations reference a Theory of Change, but these are in fact – in all cases – Logical Frameworks.

<sup>3</sup> The IOS Evaluation Office explicitly differentiates between a Logical Framework and a Theory of Change. A Logical Framework provides a reasonable, defensible and sequential order from inputs, through activities to outcomes and impacts. Logical Frameworks are descriptive. A Theory of Change on the other hand

19. This is not to say that UNESCO does not score its successes. As one evaluation report states: *'Thematically more focused interventions, such as those aimed at literacy and teacher education, are contributing to or are likely to contribute to their objectives'*.

### 3.4 Efficiency

20. Efficiency 'measures the outputs – qualitative and quantitative – in relation to the inputs. It is an economic term which is used to assess the extent to which aid uses the least costly resources possible in order to achieve the desired results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted' (OECD n.d.).

21. This is a notoriously difficult evaluation criterion and – in practice – for most evaluands impossible to assess as comparators are either not available or detailed financial information on the comparators cannot be accessed. The IOS Evaluation Office normally does not require a comparison of cost-effectiveness and none of the corporate or decentralized evaluations have done so. Instead, the evaluation reports reflect under this heading on governance, organization, staffing, program management, implementation, timely delivery and the prudent use of funds.

22. The decentralized evaluations generally provide a positive assessment, noting professional UNESCO staff and implementation partners, good coordination, and the prudent use of resources. The few exceptions raise the same themes, albeit where they did not reach the same standard. The corporate evaluations – none of which score well on efficiency – additionally point towards (i) dispersion of resources over too many activities; (ii) the understaffing of the global secretariats (often with only one professional staff position); (iii) the myriad of prevailing structures, committees, and funds whereby it is often not clear what the respective additions are; and (iv) the underutilization of UNESCO's (in principle many) partnerships.

explicitates the (implicit) assumptions regarding changes in behavior and actions of key stakeholders which *prelude and are responsible for* the achievement of the envisaged outcomes and impacts as reflected in the program's Logical Framework. A Theory of Change explicitly states *how* a program causes the intended or observed outcomes and impacts. A Theory of Change is explanatory and predictive. (M. Q. Patton 2002, Leeuw 2003) Within UNESCO, Theories of Change are also referred to a *intervention logics*.

### 3.5 Sustainability

23. Sustainability concerns *'whether the benefits of an activity are likely to continue after donor funding has been withdrawn'* (OECD n.d.). None of the evaluation reports assert that sustainability is secured without further stewardship. All evaluators point towards the importance of continued buy-in and ownership from national actors. Only one evaluation – on the Revitalization of Novobërdë/Novo Brdo Fortress – concludes that such buy-in and ownership is secured.

24. Other evaluation reports question (i) the short-term, activity-based nature of many UNESCO interventions, which prevent the build-up of sufficient local or national capacity; (ii) the insecure funding lines; and – related – (iii) the true buy-in and ownership of (local) governments. In section 3.2, we concluded that the evaluands are aligned to national goals and strategies. Why then is – more often than not – funding not secured? This may have to do with the relative importance of the evaluands within the political domain. One evaluation report puts this succinctly:

*'System resources are so limited that teachers are not being paid. Without addressing this issue, [any UNESCO intervention] will face serious pressure'.*

In other words, if the government does not have the money for basic services, it will not have the resources to spend on interventions which are – unfortunately – deemed of lesser importance.

### 3.6 Conclusion

25. This chapter asked how UNESCO performs against the standard evaluation criteria. The answer is *'mixed'*. UNESCO's work is relevant, and the organization delivers on its promises concerning activities and outputs. It struggles however in measuring its outcomes. The constrained human and financial resources amongst its implementation partners hamper the sustainability of UNESCO efforts.

26. The story on efficiency is the same. UNESCO has professional staff which (mostly) know how to work with partner organizations and within time and budget. But limited human and financial resources also constrain UNESCO's work.

### 3.7 Recommendations

27. Looking forward, the IOS Evaluation Office recommends the program sectors, field offices and category 1 institutes:

1. **Focus and select programming.** Resources within UNESCO will remain constrained for the foreseeable future. To increase its effectiveness, efficiency and sustainability, program sectors, field offices and category 1 institutes should concentrate their human and financial resources in a limited set of interventions. This will allow them to dedicate sufficient resources to design, implement and monitor these interventions and achieve a scale to allow for long-term results.
2. **Select intervention areas where national-level stakeholders can lead, both in the short and long term.** To ensure sustainability, UNESCO's work needs to be embedded in national-level agendas and structures. This implies that – even more than is the case already – UNESCO needs to align its work to the needs and priorities of local (government or non-governmental) actors. And the local actors not only have to *state* their priorities, but also be able *to show* a track record of engagement (or provide convincing evidence that they have sufficient, long-term human and financial resources in place).
3. **Enhance the evaluative analysis of UNESCO's work.** Both for accountability and learning purposes, program sectors, field offices and category 1 institutes need to be able to answer the quintessential evaluation question: what works, what does not and why? This can be done through either ex-ante evaluations which evidence the existing knowledge base on the relevance and effectiveness of an envisaged intervention area or through mid-term or ex-post program evaluations.

## 4. Quality of evaluations

*What is the quality of UNESCO's evaluation reports when measured against the UNEG Quality Checklist for Evaluation Reports?*

28. This chapter measures the 5 corporate and 12 decentralized evaluation reports against the UNEG Quality Checklist for Evaluation Reports (UNEG 2010). The authors of these evaluation reports did not receive the UNEG Checklist beforehand, nor did their Terms of Reference require adherence to this code. This makes, on the one hand, this ex-post comparison unfair. On the other hand, the checklist represents an industry standard. One would expect a reasonable level of alignment irrespective of whether the authors received the checklist from UNESCO. More important, this review is not concerned about the *individual* evaluation reports, but rather about drawing lessons from the current evaluation practice to *improve future evaluation work*. This part of the

### *Textbox 2. The applicability of the UNEG Quality Checklist of Evaluation Reports*

UNESCO evaluations – both the corporate and decentralized ones – are conducted with limited resources. Most evaluations are conducted within a US\$ 15,000 and US\$ 30,000 price range. The question is to what extent the full brunt of the UNEG Quality Checklist of Evaluation Reports can be brought to bear on the evaluation reports. On the one hand, this would be unfair. Reconstructing the Theory of Change, applying a full range of data collection methods, analyzing the data, engaging with the IOS Evaluation Office and the program sectors on the findings, and writing a comprehensive report take time, which is – given the budget – in short supply.

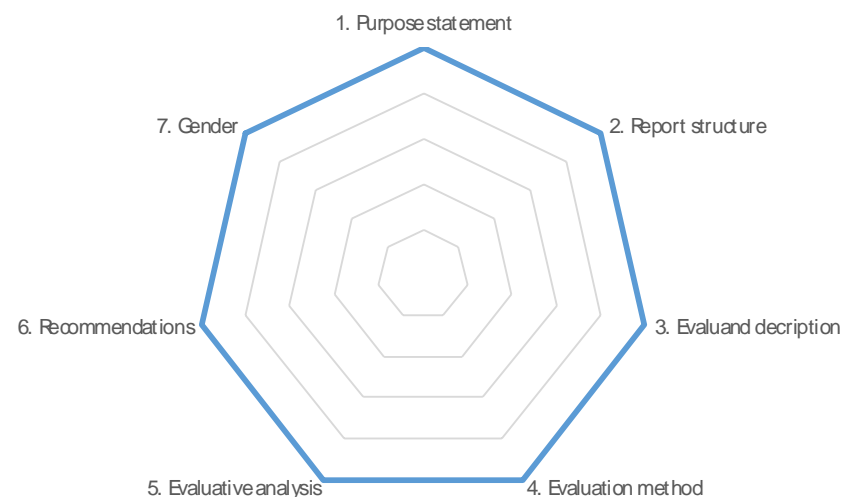
On the other hand, the UNEG Norms and Standards are to a large extent about creating clarity about what one has done and found. This standard holds irrespective of the size of the evaluation. This means, for example, that an evaluation should provide clarity on the evaluation methodology by writing out the purpose, scope, limitations and tools applied for each data collection and analysis method.

We acknowledge both of the above points. Reports are judged to meet a standard sufficiently – and thus assigned a green light – if and when the report delivers what can be expected under a US\$ 30,000 or less assignment. This reflects a judgement call on the part of the IOS Evaluation Office. At the same time, this review upholds the norm for the reports to be clear on purpose, approach, findings, conclusions and recommendations.

review is for the IOS Evaluation Office to learn and determine how it can improve its conduct, guidance or support to corporate and decentralized evaluations.

29. The checklist differentiates 8 dimensions or standards of a quality evaluation report. We deviate from the checklist on two counts. First, we combine the original dimensions 'findings' and 'conclusions' under the header 'evaluative analysis'. Second, we disregard the *human rights dimension* as UNESCO's projects and programs are human rights

Figure 7. The 7 dimensions of a quality evaluation report



based by intent and design. Figure 7 shows the 7 dimensions as applied in this study.

30. Table 2 on the next page scores each evaluation report on these 6 dimensions (the 7<sup>th</sup> dimension on gender has been separately examined within the UN-SWAP initiative of UN Women. It is reported on separately). A close look reveals two systemic issues. The evaluation reports score better on the *framing* of the evaluation and evaluation (dimensions 1 – 4) than on the evaluative analyses or recommendations (dimension 5 – 6), while gender coverage by the evaluations is uneven, especially for decentralized evaluations. We dive deeper into these observations in the subsequent sections in which we discuss the 6 quality dimensions respectively.

Table 2. Quality assessment of UNESCO's evaluation reports

		1	2	3	4	5	6
		Purpose statement	Structure	Description evaluand	Evaluation method	Evaluative analysis	Recs.
#	Corporate evaluations						
1	International Convention against Doping in Sport						
2	Bioethics and Ethics of Science and Technology Programme						
3	Work in Capacity Building in the Basic Sciences and Engineering						
4	Programme Interventions on Girls' and Women's Education						
5	Science Report						
Decentralized evaluations							
1	'Revitalizing Adult and Youth Literacy' Project						
2	Program for Palestinian University Students under Conditions of Severe Poverty						
3	Revitalization of Novobërdë/Novo Brdo Fortress Project						
4	Mahatma Gandhi Institute of Education for Peace and Sustainable Development						
5	International Fund for Cultural Diversity						
6	Two education projects: 'Expanding secondary education for Syrian refugees' and 'Bridging the gaps in secondary schooling for girls of IDP communities'						
7	'Afghanistan Heritage and Extractive Industries Development Initiative'						
8	Project 'Safeguarding Intangible Cultural Heritage through the Strengthening of National Capacities in Asia and the Pacific'						
9	Project 'Retaining Girls in Lower Secondary Schools and increasing their Learning Outcomes'						
10	Education Projects in Iraq						
11	International Programme for a Culture of Peace and Dialogue						
12	Project 'Ensuring the sustainability of the Centre for Restoration of Islamic Manuscripts in Jerusalem'						

## 4.1 Purpose statement

31. Clearly, an evaluation report needs to state the purpose of the underlying evaluation. This is to include: *'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 main evaluation questions ... and an explanation of the chosen evaluation criteria, performance standards or other criteria [to judge the evaluand against]'* (UNEG 2010).

32. Except one, all evaluation reports include a purpose statement in principle. These tend to be formulated generically and in abstract terms, such as: *'to assess the relevance and performance [of the evaluand]'* or *'to inform the Organization's decision-making and provide evidence-based recommendations on how to strengthen [UNESCO's work]'*. Some evaluation reports subsequently state whether the evaluation was mostly summative or formative in nature. Figure 6 on page 9 of the previous chapter showed that most reports include some or all of the standard evaluation criteria.

33. Most evaluation reports do however not identify the *intended users* of the evaluation, *why* they needed the evaluation, *what* their information needs were (i.e. list the users' evaluation questions over and above those defined under the standard evaluation criteria), and *how* the evaluation results will be used (in what decision-making process will they flow specifically?). (Partial) exceptions are the reports on (i) the Evaluation of UNESCO's International Convention against Doping which states that the purpose of the evaluation is to: *'inform the deliberations of the 6<sup>th</sup> session of the Conference of Parties ... and act as a catalyst for dialogue and debate about the future direction for the implementation of the Convention'*; or (ii) two education projects in Iraq for refugees and girls from internally displaced communities respectively which was conducted to: *'inform which delivery mechanisms work best under fragile circumstances'*.<sup>4</sup>

34. This review shows a clear need to strengthen the user orientation of the evaluations and – consequently – to design the evaluations around the pertinent questions of these

<sup>4</sup> The evaluation report on the project *'Ensuring the sustainability of the Centre for Restoration of Islamic Manuscripts in Jerusalem'* contains specific evaluation questions. This report is however a bit of an outlier

users about the evaluand. Such a deliberate design, together with the active engagement of the intended users in the evaluation process (for example through periodic debriefings), will go a long way in increasing the value to and use of evaluations by UNESCO's program sectors, field offices and category 1 institutes.

## 4.2 Report structure

35. Reports are to be *'logically structured with clarity and coherence'* (UNEG 2010). All evaluation reports fulfill this criterion. The checklist however also suggests information which should be entailed in the opening pages, the executive summary and the appendices of the report (see Table 3). Many evaluations cover most of these *basic data requirements* (at least in one place or another in the report). The checklist offers a

Table 3. Basic information requirements of evaluation reports

Title & openings pages	Executive summary	Appendices
<ul style="list-style-type: none"> <li>– Name of the evaluand</li> <li>– Timeframe of the evaluation</li> <li>– Date of the report</li> <li>– Location of the evaluand</li> <li>– Name(s), organization and affiliation of the evaluators</li> <li>– Name of the entity and evaluation manager commissioning the evaluation</li> <li>– Table of content, figures, tables and appendices</li> <li>– List of acronyms</li> </ul>	<ul style="list-style-type: none"> <li>– Evaluation objectives and intended users</li> <li>– Evaluation methodology</li> <li>– Key findings and conclusions</li> <li>– Recommendations</li> </ul>	<ul style="list-style-type: none"> <li>– Terms of Reference</li> <li>– List of persons interviewed</li> <li>– List of documents consulted</li> <li>– More details on methods</li> <li>– Evaluators' biodata</li> <li>– Evaluation design matrix</li> <li>– Results Framework</li> </ul>

as it constitutes a self-evaluation by the UNESCO project officer with an external validation by two thematic experts. The two validation reports read more like expert opinions than evaluations.



more structured way of presenting this information which makes this information more accessible. We have therefore changed our evaluation report template to capture these basic information requirements as suggested by the checklist and will promote the template's systematic use by external evaluators of both corporate and decentralized evaluations.

### 4.3 Evaluand description

36. Evaluations require a clear and comprehensive description of the evaluand, including the program/project's purpose, background, size, (geographical) scope, components, duration, key stakeholders, governance, organization and financial resources. All evaluations meet this criterion. The challenge is brevity. In our experience, most programs and projects let themselves be summarized within two pages which can be included in the introduction chapter. Most evaluations dedicate however a full chapter to the evaluand description with some running over 10 pages or more.

37. The checklist also suggests including the socio-political, economic, demographic and institutional context within which the evaluand operates. This is seldom done, although contextual factors do emerge when the evaluators distill the evaluations' main findings. It would nonetheless be useful to already present the key contextual factors of an evaluand at the outset of the evaluation report.

### 4.4 Evaluation method

38. The checklist presents multiple requirements for the methodology section of evaluation reports. For this review, we slimmed this list down to five key criteria: purposeful design, clear description of data collection methods *and* sources, triangulation of data sources and methods, and explicit mentioning of any limitations to the conduct of the evaluation.

39. Most evaluations score well on the description of the data collection methods and identifying their data sources<sup>5</sup>. Half of the evaluations include a section or reference to limitations in the conduct of the evaluation. Few evaluations however explain the

design of the evaluation and how the different data sets and methods provide for a representative picture of the evaluand. For example, only three evaluation reports include an Evaluation Design Matrix, namely the corporate evaluation on the UNESCO's Programme Interventions on Girls' and Women's Education and the decentralized evaluations of (i) education projects for refugees and girls from internally displaced communities in Iraq and (ii) the International Program for a Culture of Peace and Dialogue. As we will push evaluators to focus more on the intended users and use of the evaluation, we will also require evaluators to apply a more purposeful design to the evaluations, including the preparation of an Evaluation Design Matrix.

### 4.5 Evaluative analysis

40. As noted above, on this criterion we slightly deviate from the checklist. The latter distinguishes findings, conclusions and recommendations, both in the proposed structure and the quality assessment of the report. In terms of structure, most evaluation reports indeed dedicate different chapters to the findings, conclusions and recommendations. In our view, this substantially weakens the logical coherence between the presented evidence, the analysis of this evidence to distill findings and conclusions, and the formulation of recommendations. This is exacerbated by the length of the average evaluation report: 42 pages for the main report and 68 pages for the total report (see Figure 5 in Chapter 2).

41. The checklist also requires that *'findings reflect systematically an appropriate analysis and interpretation of data ... conclusions which are well-substantiated by evidence and are logically connected to the evaluation findings ... and reflect reasonable evaluative judgements relating to the key evaluation questions'*. The results on this score are mixed. Whereas five evaluation reports receive a green and two evaluation reports a red light, most evaluation reports are in the middle and struggle to fully meet this criterion.

42. We observe that most evaluation reports (i) do not distinguish between data and findings with the latter being initial insights emerging or distracted from the collected

---

<sup>5</sup> Most evaluation reports include the lists of key informants, survey respondents and consulted documents in the appendices.

data<sup>6</sup>; (ii) do not structure their arguments based on a summary presentation of the key evidence (i.e. data), an analysis of this data (i.e. findings) and a comparison of the findings to the benchmark or evaluation criteria (i.e. conclusion) – in other words, the evaluation reports are highly descriptive in nature and lack a deductive analysis of the data to make sense of the evidence and deduct answers to the evaluation questions at hand; and as noted above (iii) findings, conclusions and recommendations are decoupled from each other.

43. Linked to our suggestion in Section 3.7 to focus more on the intended users and use of an evaluation, we propose an alternative approach to and structure of the (content chapters of) the evaluation report. Each chapter should answer an explicit key evaluation question of the intended users based on a clear, coherent and fully spelled out evaluative analysis and judgment (clearly linking evidence, findings, benchmarks, conclusions and recommendations). We detail this recommendation in Section 4.8 and Annex **Error! Reference source not found. Error! Reference source not found.**

#### 4.6 Report recommendations

44. The Checklist requires recommendations to be *'firmly based on evidence and conclusions, relevant to the object and purposes of the evaluation [and] clearly identify the target group for each recommendation'*. All evaluation reports include recommendations which are relevant to the evaluand. In the previous section, we already concluded that the evidence-base is not always clear (which is not to say that they are not evidence-based). Only six evaluations include the target group for the recommendations in the recommendation section.

#### 4.7 Conclusion

45. UNESCO's evaluation reports, whether on corporate or decentralized evaluations, are *'quite good'*, but tend to be too descriptive, too long and too little user-focused. There thus remains ample room for improvement. On the one hand, making those improvements will be challenging as human and financial resources are scarce.

---

<sup>6</sup> The decentralized evaluation of the International Programme for a Culture of Peace and Dialogue is a notable exception. In the content chapters of the main report, it states the key findings and subsequently undergirds this with a reflection on key evidence (data).

Collecting a broad evidence-base, analyzing the data and answering key evaluation questions in a concise, albeit analytical way take time and thus require resources. On the other hand, many of the improvements simply require a structured approach and clear writing on what has been done.

#### 4.8 Recommendations

46. To further improve UNESCO's evaluation reports, these reports need to:

1. **Focus on use.** Behind each evaluation stands a small, core group of people who will use the evaluation results. It is the evaluators' task to serve this group by soliciting their primary questions about the evaluand, understanding the context in which they operate and the type of information they need, including them in the evaluation process through consultations and debriefing, and – ultimately – answering the primary questions in the evaluation report. Focus on usage will also allow more deliberate and purposeful evaluation designs.
2. **Become more analytical.** This is less difficult than it sounds. It means answering the prime evaluation questions of the intended users of the evaluation in an explicit and an argumentatively sound way. This means stating the primary evidence-base (i.e. the relevant data), making sense out of these data (i.e. the findings) and comparing these findings with the relevant benchmark / evaluation criteria. By focusing on answering the questions, evaluation reports can forego much of the descriptive information currently included in evaluation reports or relegate it to the appendices (for example when it concerns field mission reports).
3. **Directly follow an answer with a recommendation.** The answer to an evaluation question should be directly followed with a recommendation (if a recommendation is due). This ensures a logical connection between a conclusion and recommendation.<sup>7</sup>

<sup>7</sup> This does not preclude a separate and dedicated listing of all recommendations together (which is ideally done in the executive summary)

47. This report's conclusion and recommendations have led the IOS Evaluation Office to prepare a new guideline for the preparation of evaluation reports. This guideline is available in a separate document.

## Appendix: Quality of the evaluation reports – a scorecard

Table 4. Detailed scoring of the evaluation reports.

Systemic issues	1	2	3	4	5		1	2	3	4	5	6	7	8	9	10	11	12
<b>Purpose statement</b>																		
– Includes <b>purpose/use of evaluation</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Identifies <b>intended users</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Includes <b>clear evaluation questions</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Includes <b>well-defined evaluation criteria</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
<b>Report structure</b>																		
– <b>Clear structure</b> through the following sequence: Executive summary, introduction, program description, evaluation method, findings, conclusions and recommendations	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– To briefly state in the <b>opening pages</b> an evaluation's: – Commissioning office – Implementation status – Geographical coverage – Timeframe of data collection and analysis – Firm affiliation of the evaluators (if any)	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– To include in the <b>executive summary</b> : – Purpose statement – Description of evaluand – Evaluation methodology – Evaluation questions	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– To include as <b>annex</b> : – Terms of reference – Evaluation design matrix – Logical / results framework – Data sources – Biodata of the evaluators	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●

Systemic issues	1	2	3	4	5		1	2	3	4	5	6	7	8	9	10	11	12
<b>Description of evaluand</b>																		
– <b>Comprehensive description</b> through purpose statement, background, size, (geographical) scope, components, duration, key stakeholders, governance, organization and financial resources.	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Inclusion of <b>Logical Framework or Theory of Change</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
<b>Evaluation methods</b>																		
– Selection of evaluation methods / <b>evaluation design</b> matrix	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Clear description of data collection <b>methods</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Clear description of data collection <b>sources</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– <b>Evidence-base</b> / triangulation	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Includes <b>limitations</b> of evaluations	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
<b>Evaluative analysis</b>																		
– Indictive and deductive <b>analysis of collected data</b> = findings	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– <b>Benchmark</b> findings against evaluation criteria = conclusions	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– <b>Answers the evaluation questions</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Findings, conclusions and recommendations <b>connected</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
<b>Recommendations</b>																		
– <b>Clear and well-substantiated</b>	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– Clear <b>to whom</b> it is directed	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
– <b>Relevant</b> to the purpose of the evaluation	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●

Legend: ● = green light; ● = orange light; ● = red light

Note: The UNEG Checklist identifies 4 – 7 sub-criteria per quality standard. For this review, we reduced the number of sub-criteria.

## References

- Leeuw, Frans L. 2003. "Reconstructing Program Theories; Methods Available and Problems to be Solved." *American Journal of Evaluation* (Elsevier Science Inc. on behalf of the American Evaluation Association) Volume 24 (Nr. 1): pg. 5-20.
- Morra Imas, Linda G. , and Ray C. Rist. 2009. *The Road to Results. Designing and Conducting Effective Development Evaluations*. Washington DC: The World Bank.
- OECD. n.d. *OECD DAC Network on Development Evaluation. Evaluating Development Co-operation. Summary of Key Norms and Standards*. 2nd edition. Paris: OECD.
- Patton, Micheal Quin. 2002. *Qualitative Research & Evaluation Methods*. 3rd edition. Thousand Oaks: Sage Publications.
- UNEG. 2010. *UNEG Quality Checklist for Evaluation Reports*. Guidance Document, New York: United Nations Evaluation Group. <http://www.unevaluation.org/document/detail/607>.
- UNESCO. 2015. *UNESCO Evaluation Policy*. Paris: UNESCO.