



United Nations
Educational, Scientific and
Cultural Organization

Internal Oversight Service
IOS/EVS/PI/148 REV.2

Synthetic Review of Evaluations in the UNESCO System

Evaluation Office

May 2016

ABSTRACT

The Evaluation Office of the Internal Oversight Service at UNESCO has commissioned a synthetic review study of existing evaluations in order to assess the coverage and quality of the evaluative work across the UNESCO system, and to identify systemic issues that constrain or enable the organization's work across the UNESCO system. The study, covering decentralized and corporate evaluation reports, adopted a 4-tier approach to the study, which included: a coverage analysis, a succinct meta-evaluation, a synthetic review of the relevance and effectiveness (for illustrative purposes, given the challenges in evaluation coverage and quality), and a synthetic review of cross-cutting issues that enable or hinder implementation and outcome achievement. Among other things, the study has found that the current evaluation coverage is fragmented and uneven, and that there are significant challenges to improving the quality of decentralized evaluations. While evaluations in principle constitute the most credible source of evidence for synthetic analysis across programmatic areas of work, the current coverage and quality of evaluations does not yet allow for such analysis. The study recommends that UNESCO should develop a clear definition of a programmatic area of work, which should be tied to a standardized information cycle, it should strengthen its decentralized evaluation system, and it should improve its data collection and tracking system of decentralized evaluation reports.

EVALUATION TEAM

Overall responsibility	Ms Susanne Fruch, Director Mr Amir Piric, Head of Evaluation
Evaluation manager	Mr Jos Vaessen, Principal Evaluation Specialist
Administrative support	Ms Arushi Malhotra, Evaluation Assistant
Consultant	Ms Estelle Raimondo

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	4
CHAPTER 1: BACKGROUND	8
1 RATIONALE	8
2 PURPOSE	9
3 SCOPE AND DELIMITATION	9
4 METHODOLOGY	11
CHAPTER 2: COVERAGE ANALYSIS.....	21
1 COVERAGE AT THE UNESCO-WIDE LEVEL.....	21
2 COMPARISON ACROSS SECTORS	23
3 COVERAGE WITHIN SECTORS	25
4 KEY FINDINGS	29
CHAPTER 3: SUCCINCT META-EVALUATION.....	30
1 BASIC QUALITY OF EVALUATIONS	30
2 BASIC QUALITY OVER TIME.....	36
3 MINIMUM REQUIREMENTS FOR SYNTHETIC ANALYSIS.....	38
4 KEY FINDINGS	40
CHAPTER 4: SYNTHETIC ANALYSIS OF EFFECTIVENESS AND RELEVANCE OF UNESCO WORK - ILLUSTRATIVE EVIDENCE FOR SELECTED EXPECTED RESULTS	41
1 ILLUSTRATIVE SYNTHETIC ANALYSIS OF RELEVANCE AND EFFECTIVENESS	41
2 THE CONTRIBUTION OF EVALUATIONS TO A COMPARATIVE PROGRAMME ASSESSMENT	59
3 KEY FINDINGS	61
CHAPTER 5. SYNTHESIS OF CROSSCUTTING ISSUES.....	63
1 CROSSCUTTING CHALLENGES.....	63
2 CROSSCUTTING ENABLING FACTORS	74
3 KEY FINDINGS	77
CHAPTER 6. OVERALL CONCLUSION AND RECOMMENDATIONS	78
ANNEX 1: LIST OF PROGRAMMATIC AREAS.....	79
ANNEX 2: ASSESSING MINIMUM REQUIREMENTS FOR SYNTHETIC ANALYSIS BY EXPECTED RESULTS .	84

LIST OF TABLES

Table 1. Recap of the sampling strategy for each part of the analysis	11
Table 2. Assessment framework for the quality of evaluation reports.....	14
Table 3. Assessment framework for minimum requirements for synthetic analysis.....	15
Table 4. Assessment framework for ‘relevance’	16
Table 5. Assessment framework for ‘effectiveness’	17
Table 6. Total number of evaluation reports by Expected Results (n = 261*)	28
Table 7. Assessment Matrices for Education Policy and Planning (n = 9).....	42
Table 8. Assessment Matrices for Capacity for water management (n = 8 reports)	45
Table 9. Assessment Matrices for Youth (n = 5 reports)	50
Table 10. Assessment Matrices for Diversity of cultural expressions (n = 8 reports)	52
Table 11. Assessment Matrices for Community and pluralistic media (n = 3 reports)	57
Table 12. Potential contribution of evaluative evidence to comparative programme assessment.....	61

LIST OF FIGURES

Figure 1. Total number of evaluation reports by type (n = 261).....	21
Figure 2. Total number of evaluation reports by timing (n = 261).....	22
Figure 3. Total number of evaluation reports by year of publication (n = 261).....	23
Figure 4. Total number of evaluation reports by sector (n = 261).....	23
Figure 5. Total number of evaluation reports by type and sector (n = 261).....	24
Figure 6. Total number of evaluation reports by type of commissioner (n = 261)	25
Figure 7. Total number of evaluation reports by level of implementation (n = 261)	25
Figure 8. Total number of evaluation reports by Expected Results (n = 261*).....	27
Figure 9. Total number of final evaluation reports by effects (n = 187)	31
Figure 10. Total number of final evaluation reports by effects and by type (n = 187).	31
Figure 11. Total number of final evaluation reports by lessons learned (n = 187)	32
Figure 12. Total number of final evaluation reports by type and by lessons learned (n = 187)	32
Figure 13. Total number of final evaluation reports by clarity of causal chain (n = 187).....	33
Figure 14. Total number of final evaluation reports by clarity of causal chain and type (n = 187).....	33
Figure 15. Total number of final evaluation reports by logframe (n = 187)	34
Figure 16. Total number of final evaluation reports by type and by quality points (n = 187)	35
Figure 17. Average quality points of final evaluation reports by sector	35
Figure 18. Total number of final evaluation reports by average quality points and by type of commissioner: "independent"= yes (n = 187)	36
Figure 19. Distribution of average quality points for the sample of final evaluation reports included in the 2013 Diagnostic study (n = 78).....	37
Figure 20. Distribution of average quality points for the sample of recent final evaluation reports (after 2013; n = 69).....	37
Figure 21. Comparison in quality between decentralized final evaluation reports from 2013 or before, and after 2013 (n = 147).....	38
Figure 22. Distribution of final evaluation reports with a quality score of 4 or more by ER (n = 157)...	39

Figure 23. Distribution of final evaluation reports with a quality score of 4 or more and commissioned	39
Figure 24. Total number of ERs that meet the minimum requirements	40
Figure 25. Education Policy and Planning: Relevance and Effectiveness (n = 9).....	48
Figure 26. Capacity for water management: Relevance and Effectiveness (n = 9).....	49
Figure 27. Youth: Relevance and Effectiveness (n = 5)	55
Figure 28. Diversity of cultural expressions: Relevance and Effectiveness (n = 8).....	56
Figure 29. Community and pluralistic media: Relevance and Effectiveness (n = 3).....	58
Figure 30. Two by two decision matrix with fictitious Expected Results	60

LIST OF BOXES

Box 1. Example of an Education Policy and Planning project with high relevance and effectiveness ...	43
Box 2. Example of an Education Policy and Planning intervention with high relevance and limited (evidence of) effectiveness.....	44
Box 3. Example of a joint Water Management intervention with high relevance and effectiveness	46
Box 4. Example of a Water Management intervention with high relevance and low effectiveness	47
Box 5. Example of a Youth joint intervention with moderately high relevance and high effectiveness	51
Box 6. Example of a Diversity of Cultural Expressions intervention with high relevance and effectiveness	53
Box 7. Example of a joint Diversity of Cultural Expressions intervention with moderately high relevance and limited effectiveness	54
Box 8. Example of a Community and Plurality of Media joint intervention with low relevance and high effectiveness	58
Box 9. Examples of challenges of working within UNESCO "families" stemming from Category I institutes evaluations	68
Box 10. Examples of challenges of working within UNESCO "families" stemming from IHP-VII evaluation.....	69

EXECUTIVE SUMMARY

Background

UNESCO has recently stepped up its efforts to improve its results-reporting framework, which constitutes an important step toward better evidence-informed decision-making. Evaluations¹ are valuable sources of evidence for complementing the self-reported evidence that inform much of UNESCO's statutory results-reporting. In the preparation of the quadrennial Strategic Results-Report (SRR) of the Organization, the different entities of the Secretariat were encouraged to consult evaluations in the development of their self-assessments. To complement this perspective, and in line with the recently endorsed UNESCO Evaluation Policy (296 EX/24.INF) the IOS Evaluation Office has commissioned a synthetic review of existing evaluations.

The literature on synthetic review and the experience of other international organizations suggest that a credible synthetic review to inform strategic decision-making relies on the following three building blocks:

- A consistent delimitation of a programmatic area of work (unit of analysis);
- A sufficient evaluation coverage of each area of work; and
- A strong evidence base for synthetic judgment on whether a particular area of work makes a difference (quality).

Previous studies (e.g. IOS/EVS/PI/128.REV; IOS/EVS/PI/136 REV.) have shown that none of these three criteria are likely to be fully met in the case of UNESCO. Consequently, part of the synthetic review focuses on how to improve the evaluative evidence base and its use in the future, and consequently strengthen the potential for evaluations to inform strategic decision-making processes in the Organization.

Purpose

The overall purpose of the synthetic review is to generate evidence from existing evaluations of UNESCO's interventions. More specifically, the study has the following purposes:

- To assess the coverage of evaluative work across the UNESCO system;
- To assess basic aspects of quality of evaluation reports, complementing the 2013 meta-evaluative study of existing evaluations (IOS/EVS/PI/128.REV);
- To generate synthetic evidence on the relevance and effectiveness of programmatic areas of work; and
- To identify systemic issues that constrain or enable UNESCO's work across the UNESCO system.

Approach

The study covers decentralized evaluation reports that were published between 2009 and September 2015, as well as corporate evaluation reports published between 2008 and September 2015. The choice of a rather long time-span was justified by the fact that IOS conducted a series of evaluations on all of UNESCO's Strategic Programme Objectives (SPOs) starting in 2008, which provide some level of coverage of all of UNESCO's programmatic areas of work as delineated in the previous Mid-Term strategy of the Organization (34 C/4). In total, the study covered 261 evaluation reports (41 corporate and 220 decentralized evaluations).

¹ Evaluations are defined here as assessments conducted by external experts. They should be clearly distinguished from final narrative reports and other assessments that are based on self-evaluation, i.e. exercises conducted by UNESCO programme staff. Decentralized evaluations are managed by the Secretariat (HQ, Field Offices, Category I Institutes). Corporate evaluations are managed by the IOS Evaluation Office.

More specifically, the study adopted a four-tiered approach:

- *A coverage analysis*, which consists of mapping evaluation reports onto UNESCO's programmatic areas (Expected Results (ERs)), and assessing the level of evaluative coverage of each programmatic area;
- *A succinct meta-evaluation*, which consists of assessing the minimum quality of evaluation reports² on a set of easy to measure quality criteria;
- *A synthetic review of the relevance and effectiveness* of one programmatic area for each of the five Major Programmes of UNESCO for which the evaluation coverage and quality meet the minimum criteria for synthetic analysis; and
- *A synthetic review of cross-cutting issues* that enable or hinder project delivery or processes of change.

As part of component 3, this study sought to illustrate what a comparative analysis of programmatic areas of work using evidence from evaluations would look like if the three above-mentioned conditions (unit of analysis, coverage, quality) were met for the majority of UNESCO's programmatic areas. UNESCO has identified the following criteria (197 EX/5 Part IV): relevance, capacity to deliver, comparative advantage, tangible results, and sustainability. In principle, evaluation reports have a comparative advantage in providing information on a number of these criteria, as illustrated in the report.

Findings

On evaluation coverage:

- Evaluative coverage is uneven across sectors: while some UNESCO sectors demonstrate a rather good evaluation coverage (Culture, Education and Natural Sciences), other sectors have a very low evaluation coverage (Social and Human Sciences, Communication and Information).
- Within sectors, evaluation coverage is very uneven across ERs: while some ERs are particularly well-covered, others are essentially "evaluation-free" with no evaluations at the level of ERs or no "within ERs" evaluation. This is the case for all UNESCO sectors.

On the quality of evaluations:

- Overall, the vast majority of evaluation reports meet the basic reporting requirements in terms of including information on implementation, output delivery, effects, and lessons learned. The evaluation reports provide abundant and rich descriptions of activities and outputs, as well as an increasingly consistent effort to draw lessons from the intervention. However, the evidence on effects (outcome or impact) is often very patchy.
- The quality of reports has been stable over time. Notably, there has been no substantial change in the quality of reports after the 2013 Diagnostic Study of Evaluations of UNESCO's Extrabudgetary Activities.
- The minimum requirements for assessing the relevance and effectiveness of a programmatic area of work on the basis of existing evaluations— based on the criteria of sufficient coverage, minimum quality, and independence—are met for only a fifth of UNESCO's 47 ERs. Consequently, comparative assessment of programmatic areas of work on the basis of existing evaluations is currently not possible.

On the potential for synthesizing evaluative evidence on the relevance and effectiveness of programmatic areas of work:

² In evaluation theory, the term meta-evaluation is often used to refer to studies that assess the quality of evaluations. Although the present study has many of the characteristics of a meta-evaluation, it covers a much larger sample than most meta-evaluative studies and as a result focuses on fewer dimensions.

- In contrast to self-evaluation and self-reporting, an external evaluation is an independent inquiry based on a systematic process of data collection and analysis. Consequently, evaluations have the potential to provide more credible evidence on a number of strategic performance issues, including outcome (expected results) achievement. Given that there is currently no evaluation strategy at the ER level, assessment of effectiveness and relevance at the programmatic area level has to be extrapolated from a rather patchy evidence base. Nevertheless, the illustrative assessment of the programmatic areas that meet the minimum requirements for synthetic review demonstrates that it is possible to distinguish underperforming from well-performing ERs.

On cross-cutting challenges and enabling factors:

- Across sectors, levels of interventions and domains of expertise, it is clear that UNESCO has a number of strengths and attractive features, which— when they come together—have enabled important programmatic successes. Chief among these are: UNESCO's participatory and interdisciplinary programming practices, its large network of institutional partners, and its potential to mobilize and deploy a critical mass of diverse expertise.
- Nevertheless, the coalescence of these important ingredients for a relevant and effective intervention seems to be the exception, rather than the rule. A number of key structural challenges—indistinctive of sectors or areas of work—have hindered UNESCO's capacity to make a difference. Chief among these are: a lack of strategic focus that affects the quality and potential for impact of its work at all levels of intervention, limited financial and human resources, issues of coordination and strategic alignment that weaken the potential of the wider UNESCO network, dispersed governance systems, and a number of operational challenges.

Overall conclusion and recommendations

Evaluations are potentially the most credible source of evidence on a number of strategic performance issues, including outcome (expected results) achievement of UNESCO's programmes. However, the current evaluation coverage of UNESCO's programmes is fragmented and uneven. In addition, significant challenges to improving the quality of decentralized evaluations remain. To strengthen the role of evaluations in supporting evidence-informed decision-making, the study recommends the following:

Recommendation 1: The Organization continues to spend too much effort on (micro) activity assessment and reporting at the cost of adequate strategic reflection and assessment at a higher programmatic level. **UNESCO should develop a clear definition of a programme or programmatic area of work** (e.g. around an ER). **Each programmatic area of work should be tied to a standardized information cycle**, constituting the basis for better planning, implementation, monitoring and evaluation, which should include the following elements:

- a strategic analysis, resulting in the development of an intervention logic of the programmatic area of work that clearly articulates the main activities, outputs (deliverables) and outcomes (expected results). This would provide the basis for:
- a results framework with clear and comprehensive indicators at output and outcome levels, which in turn constitutes the basis for:
- continuous monitoring of programme output delivery and outcome achievement, and periodic evaluation of the programme.

The standardization of the unit of analysis (a programme) for planning and information collection purposes, in combination with a harmonized approach to information collection across programmes, closely resembles the project cycle principle that has been successfully adopted in many organizations across the globe. It has the potential to both improve the quality of planning, monitoring and

evaluation, as well as lowering the transaction costs for doing so, enhancing UNESCO's potential to become more efficient and effective at the same time.

The framework described above represents a deepening of a process that has already been set in motion through various reform efforts and Executive Board decisions and would strengthen the foundation for supporting the Organization's reform toward improved:

- results-based budgeting;
- evidence-informed decision-making on the strategic allocation of financial resources and the identification of strategic priorities.

Recommendation 2: To improve the role of evaluation to support evidence-informed decision-making, **UNESCO should strengthen its decentralized evaluation system**. More particularly, to improve the quality and coverage of evaluations, the Organization should:

- increase the resources available for evaluation through improved budgeting practices and procedures for extrabudgetary activities;
- strengthen staff capacities for managing decentralized evaluations;
- strengthen the mechanisms for planning, backstopping and information exchange of decentralized evaluations.

Recommendation 3: **UNESCO** (i.e. BSP, IOS and KMI) **should improve the data collection and tracking system of decentralized evaluation reports**. This would allow the Organization to improve its database of decentralized evaluation reports, and consequently the quality and use of periodic meta-evaluations and syntheses of evaluation reports.

CHAPTER 1: BACKGROUND

1| RATIONALE

The rationale for a synthetic review of evaluative evidence

In the past few years a number of policy documents and findings from corporate evaluations have justified the need for periodic synthetic reviews of evaluative evidence across the UNESCO system. Most notably:

- The Evaluation Policy (296 EX/24.INF) endorsed by UNESCO's Executive Board in March 2015 mandates the IOS Evaluation Office to conduct periodic meta-evaluations to assess the quality of decentralized evaluations as well as synthetic reviews of the content of these evaluations. In time, these should feed into the Organization's Strategic Results Report.
- The 2013 Diagnostic Study of Evaluations of UNESCO's Extrabudgetary Activities (IOS/EVS/PI/128.REV) highlighted a number of problems with the quality of the evaluation reports, and concluded on the importance of continuing to periodically monitor their quality.
- The 2014 Formative Evaluation of UNESCO's Results-Reporting (IOS/EVS/PI/136 REV.) identified clear weaknesses in UNESCO's results-reporting which led to a number of structural reforms. The Executive Board endorsed (195 EX/Decisions para7) the principle that a clear distinction should be made between reporting on activities and output delivery (primarily through programme implementation reports), and reporting on ERs and outcomes (with a role for evaluations to substantiate such reporting).
- Moreover, a survey conducted in the framework of the evaluation of UNESCO's Results-Reporting, highlighted the expectations of member states that the Secretariat should more strategically analyze the challenges in the implementation of UNESCO's programmes, and the need for more synthesized and aggregated results information to be presented at the level of UNESCO's different areas of work. There was also an expectation of a clearer distinction between outputs and outcomes in the reporting.
- While in principle evaluations have a comparative advantage to generate outcome-related information that can be integrated into decision-making processes, it was highlighted in the 2013 Diagnostic Study and the 2014 Results-Reporting Evaluation that UNESCO does not yet have a comprehensive evaluation evidence base to substantiate the quadrennial Strategic Results Report. Consequently, the SRR builds on multiple sources of information and this synthetic review as well as its underlying evaluations constitute a complementary source of information.

Towards an evidence-informed decision making process at UNESCO

In line with the principles of Results-Based Management and Results-Based Budgeting, the idea behind a synthetic review of evaluative evidence is to present a comprehensive analysis of UNESCO programmes (including the relevance, comparative advantages, and effectiveness of UNESCO's work) with a view to providing comparative data to the Governing Bodies to support decision-making on strategic directions and the allocation of human and financial resources of the Organization, notably at the level of UNESCO's ERs.

Currently, the evidence-base is too weak and fragmented to inform decision-making in a comprehensive manner. However, in the future there is potential for strengthening the decision-making criteria and their underlying evidence base. In this study, we illustrate what such a report would look like, if the three following conditions were systematically met:

1. A consistent delimitation of a programmatic area of work (unit of analysis);
2. A sufficient evaluation coverage of each area of work; and
3. A strong evidence base for synthetic judgment on whether a particular area of work makes a difference (quality).

2| PURPOSE

The overall purpose of the synthetic review is to generate evidence from existing evaluations of UNESCO's interventions. More specifically, the study has the following purposes:

- To assess the coverage of evaluative work across the UNESCO system;
- To assess basic aspects of quality of evaluation reports, complementing the 2013 meta-evaluative study of existing evaluations (IOS/EVS/PI/128.REV);
- To generate synthetic evidence on the relevance and effectiveness of programmatic areas of work; and
- To identify systemic issues that constrain or enable UNESCO's work across the UNESCO system.

3| SCOPE AND DELIMITATION

The study includes evaluations of interventions, which (at least in part) are funded by or implemented by UNESCO.

1. Time-span

The study covers decentralized evaluation reports³ that were published between 2009 and September 2015, as well as corporate evaluation reports that were published between 2008 and September 2015. The choice of a rather long time-span was justified by the fact that IOS/EVS conducted a series of evaluations on all of UNESCO's Strategic Programme Objectives (SPOs) starting in 2008, thereby offering some minimum coverage of all of UNESCO's programmatic areas as delineated in the Organization's Mid-Term strategy at that time (34 C/4). In addition, in 2010 an Independent External Evaluation of UNESCO was conducted, offering a comprehensive coverage of the systemic challenges and opportunities for the institution. Notwithstanding the foregoing, this study particularly focuses on the period 2012-2015, which corresponds to UNESCO's prior Programme and Budget (36C/5: 2012-2014) and the beginning of the current Programme and Budget (37C/5: 2014-2017).

2. Type of information

The present exercise encompasses external (not self-) evaluations of UNESCO's activities only. To avoid confusion in terminology, we define external evaluation reports as follows: an evaluation conducted by entities and/or individuals outside the team in charge of the operations. The exercise focuses on all evaluations managed or conducted by UNESCO or managed by other organizations for projects/activities in which UNESCO has played a role.⁴

³ Decentralized evaluations are managed by the Secretariat (HQ, Field Offices, Category I Institutes). Corporate evaluations are managed by the IOS Evaluation Office.

⁴ Evaluations covering interventions in which UNESCO's role was (close) to negligible were excluded (e.g., in a bilateral donor's evaluation of ten years of development cooperation, UNESCO may be mentioned as one of many of implementing partners or organizations receiving funding).

A coverage analysis is undertaken to look at the distribution of evaluations across UNESCO's areas of work and covers both mid-term and final evaluation reports. However, the other parts of the study focus exclusively on final evaluation reports. Both corporate (commissioned by IOS) and decentralized evaluations (commissioned by UNESCO's sectors, Field Offices, or partners) are taken into account.

The methodology section below explains the purposive sampling strategy for each part of the analysis, which applies the following overarching principles:

- On the one hand, to assess evaluative coverage and to conduct a meta-evaluation, what matters is the total number of evaluations (and their distribution);
- On the other hand, to synthesize information on relevance and effectiveness, what matters is the quality of the evidence.

3. Type of evaluand

The evaluation reports included in the study cover interventions that were funded both through regular programme budget and extra-budgetary resources. No discrimination was made in terms of the level of intervention, nature of the evaluand, or topic addressed.

However, a clear distinction was made between two types of evidence:

- The evidence base for the purpose of synthesizing information on relevance and effectiveness of UNESCO's programmatic area, which stems from evaluation reports for which the evaluand is a project, programme or cluster of programmes; and
- The evidence base for the purpose of addressing cross-cutting issues (enabling and hindering factors), which stems from evaluation reports for which the evaluand can also be an administrative unit (e.g., Institute or Field Office), process (e.g., results-reporting), delivery mechanism (e.g., prize or anniversary), or strategy (e.g., Priority Africa or Priority Gender Equality).

4. Programmatic areas

To conduct a full-fledged cross-sectoral comparison of results achieved, the following conditions need to be met:

- A consistent delimitation of a programmatic area of work (unit of analysis);
- A sufficient evaluation coverage of each area of work; and
- A strong evidence base for synthetic judgment on whether a particular area of work makes a difference (quality).

Since these conditions were not met for the majority of ERs, this study devises a synthesis methodology and applies the methodology in an illustrative manner to one Expected Result per sector, for which the above-mentioned conditions are met.

5. Integration of information on Priority Gender Equality and Priority Africa

It was not possible to systematically harvest information on UNESCO's achievements under Priority Gender Equality from evaluation reports for the following reasons: (i) gender is not mainstreamed systematically in interventions, nor is it systematically tracked into evaluation reports; and (ii) there were a few examples of gender-specific interventions in the sample of evaluation reports, but these examples were not representative of UNESCO's work in gender equality or women's empowerment to provide a meaningful picture of UNESCO's achievement under Priority Gender Equality. Moreover, there was already a recent and rather comprehensive review of UNESCO's Priority Gender Equality,

which specifically assessed UNESCO's gender-specific initiatives and the organization's progress towards gender mainstreaming, with a number of important findings (IOS/EVS/PI/125. REV2).

With regards to Priority Africa, it is even more difficult to synthesize information from evaluation reports on the achievement of UNESCO under "Priority Africa", due to the absence of a clear theory of change for the priority, and clearly defined focus areas under the priority, as highlighted in the 2012 corporate evaluation of UNESCO Priority Africa (IOS/EVS/PI/117). Moreover, two important corporate evaluations have been conducted on the topic: the aforementioned Evaluation of UNESCO Priority Africa (2012) and the 2015 review of UNESCO's field reform in Africa (IOS/EVS/PI/143 REV.), with important conclusions.

Consequently, given the existence of recent synthetic report, in this review we do not explicitly synthesize evidence on Priority Africa and Priority Gender Equality. Nevertheless, we do take into account the evaluation reports in our analysis of cross-cutting issues (Chapter 5).

4| METHODOLOGY

The objectives of this study are addressed through a four-tiered approach:

1. *A coverage analysis*, which consists of mapping evaluation reports onto UNESCO's programmatic area and assessing the level of evaluative coverage of each programmatic area;
2. *A succinct meta-evaluation*, which consists of assessing the quality of evaluation reports⁵ on a range of simple quality criteria;
3. *A synthetic review of relevance and effectiveness* of one programmatic area per core sector for which the evaluation coverage meets the minimum criteria for synthetic analysis; and
4. *A synthetic review of cross-cutting issues* that enable or hinder project delivery or processes of change.

1. Data collection

The study relied on a centralized database hosted by UNESCO IOS that was put in place in 2013 for the first *Diagnostic Study of Evaluations of UNESCO's Extrabudgetary Activities*. To update the database and to include corporate evaluation reports (which were not part of the Diagnostic Study) additional data collection procedures were designed and implemented:

- Inclusion of corporate evaluation reports (available on UNESCO-IOS website);
- Formal requests to Sectors, Field Offices and Category I Institutes for submitting decentralized evaluation reports to IOS, including two follow-up mailings.

Basic information was recorded in a database in excel. In addition, all the reports collected were stored in a database in MaxQDA, a content analysis software package, which was used to perform all the major steps of the analysis.

2. Data Analysis

What follows is a description of the methodological principles and sampling strategies applied to each of the four parts of the study. Table 1 recapitulates the various samples.

Table 1. Recap of the sampling strategy for each part of the analysis

⁵ In evaluation theory, the term meta-evaluation is often used to refer to studies that assess the quality of evaluations. Although the present study has many of the characteristics of a meta-evaluation, it covers a much larger sample and as a result is therefore less detailed than a meta-evaluation.

Part of the analysis	Sampling criteria	Sample size
1. Coverage Analysis	Full population of reports	n = 261
2. Succinct Meta-Evaluation	Final evaluation reports only	n = 187
3. Synthesis of relevance and effectiveness	Final evaluation reports for 5 ERs meeting the minimum requirements for synthetic analysis	n = 49
4. Synthesis of cross-cutting challenges and enabling factors	In addition to the sample for (3), all other corporate evaluation reports dealing with cross-cutting issues	n = 68

a. Methodological principles for the coverage analysis

The entire universe of external evaluation reports that IOS was able to identify until September 2015 amounted to 296 reports, including both corporate and decentralized evaluation reports. 267 reports to be covered in the analysis were transferred by IOS to the consultant. This number excludes reports that were written in multiple languages, or multiple reports on the same intervention (e.g., self-assessments), or reports that were included in the database but were outside the scope of the synthetic review (e.g., decentralized reports before 2009). Six reports were ultimately deleted from the list because of duplications. The final sample thus counts 261 reports.

Selecting the right unit of analysis (at what level should we assess UNESCO interventions?) is an important challenge. Given UNESCO's diverse portfolio of activities, not only in terms of theme but also in terms of the magnitude of funding and the intervention modality, the choice of a unit of analysis for synthesis and comparison is not simple. The lowest level of funding within the UNESCO system is the activity. However, this level of analysis is too narrow to permit a strategic perspective (e.g., the way particular activities fit together). On the other hand, considering the Main Line of Action (MLA) as the primary unit of analysis and comparison would negatively affect the quality of the analysis given the heterogeneity of underlying interventions at this very aggregate level.

Given the disadvantages of these two units of analysis, this study retains an intermediary level of analysis (between the activity and the MLA), which is the ER. As a grouping, a given ER still encompasses a substantial heterogeneity in activities, but the level for coherence is higher than at the MLA level. Moreover, it is an important reference for decision-making purposes. For example, the ER level is pertinent in UNESCO's Results-Based Budgeting processes.

The reference document for mapping the evaluations is UNESCO's Programme and Budget for the period 2014-2017, the 37 C/5.⁶

The mapping of evaluation reports onto ERs relied on the following assumptions:

- There is consistency over time in the formulation of ERs: interventions that were evaluated between 2008-2015 and were, thus, designed and implemented within or before this period contributed to ERs that were closely related to the ERs of the 37 C/5;⁷
- Extrabudgetary activities can also be mapped onto ERs contained in the 37C/5;
- Evaluation reports cover interventions that contribute primarily to ERs within a single sector of intervention; or,
- Evaluation reports can cover interventions that contribute to multiple ERs within a single sector:
 - For evaluation reports that relate to a more aggregate level of analysis (e.g., SPOs) where it was possible to identify findings at the level of the ER, the reports were mapped onto

⁶ The full list of Expected Results by Sector is presented in Annex 1.

⁷ In some sectors this assumption is not clearly met. For example, in the SHS sector, a number of programmatic areas that featured in previous C5 documents (e.g. peace, gender equality), are missing in the 37 C/5.

multiple ERs. As a result, a single evaluation report can be counted multiple times in the analysis because it contributes to multiple ERs.

- When the findings cannot be categorized by ERs (e.g., for evaluation of global priorities or evaluations of corporate processes or units), the report was categorized under "Miscellaneous" and used primarily for the analysis of cross-cutting issues.

The evaluation coverage—in terms of the number of evaluation reports— was assessed at three levels:

- At the level of the Organization;
- Across sectors; and
- Within sectors, across ERs.

b. Methodological principles for the succinct meta-evaluation

The succinct meta-evaluation was based on final evaluation reports (n = 187) only (excluding, mid-term evaluations and 'other' evaluations), with two primary objectives:

- Assessing the quality of the reports (including possible changes since the 2013 Diagnostic Study); and
- Determining whether or not the evaluative coverage of a given ER meets the minimum criteria for synthetic analysis.

As a basis for a descriptive statistical analysis of the quality of the collected reports five variables were defined focusing on the content of the report. For the sake of intertemporal comparison, the quality criteria from the 2013 Diagnostic Study were retained. The matrix is presented in Table 2. For further information on the methodology see IOS (2013).

Table 2. Assessment framework for the quality of evaluation reports

Criteria	Definition	Scale
Implementation	Whether or not a report contains information on project implementation	Dichotomous variable: yes versus no
Outputs	Whether or not a report contains information on project outputs (those elements that are directly generated/delivered by the project)	Dichotomous variable: yes versus no There is often a lack of clarity on the exact wording for output. Terms like outputs, deliverables, results, or even outcomes may be (correctly or wrongly) used to refer to outputs.
Effects	Whether or not a report contains information on project effects (intended or unintended change due directly or indirectly to an intervention)	Dichotomous variable: yes versus no There is often a lack of clarity on the exact wording for effect. Terms like outputs, deliverables, results or even outcomes may be (correctly or wrongly) used to refer to effects.
Lessons learned /Challenges	Whether or not a report contains information on lessons learned	Dichotomous variable: yes versus no Lessons highlight strengths (assets) or weaknesses (challenges) in preparation, design, and implementation that affect performance, output, outcomes and impact.
Clarity of causal chain	Whether or not the causal language and logic in a report are clear	Dichotomous variable: yes versus no This variable assesses: whether or not the distinctions between implementation, outputs and effects are clear and consistent or not. Reports that only contain information on implementation and outputs (but not on effects) are coded 'no'.
Quality	Composite variable, proxy of quality of reports summarizing information on variables: - implementation - outputs - effects - lessons learned - clarity of causal chain	Ordinal variable ranging from 0 to 5

For most variables we either compared quality by sector or by period, taking 2013 (the date of the Diagnostic Study as the cut-off year for before and after comparison).

c. Methodological principles for inclusion of reports in the synthetic analysis of relevance and effectiveness

For the analysis of evaluation coverage the primary factor of importance is the number of evaluation reports and its distribution. By contrast, for the synthetic analysis of evidence on relevance and effectiveness, other factors than quantity (coverage) need to be taken into account, including the quality and credibility of the evaluative evidence underlying the findings and recommendations of evaluation reports. A thorough assessment of the quality of evaluations would involve an in-depth review of each report, which is unfeasible for the large number of reports included in this study. Consequently, we relied on the criteria included in the succinct meta-evaluation analysis laid out above as a proxy for quality. The credibility of evaluative evidence is also a criterion that is difficult to assess. The level of independence of the entity commissioning the evaluation reports is often used as a proxy for the credibility of the evaluative analysis (e.g., UNEG Norms and Standards and ECG Big

Book on Good Practice Standards). In this analysis we use the variable of whether or not a report was commissioned by a central evaluation office (whether at UNESCO or not) as a proxy for evaluative independence.

Determining the minimum requirements for synthetic analysis (of relevance and effectiveness) is a somewhat arbitrary exercise which ideally should include some type of decision rule regarding the three criteria discussed above (coverage, quality and independence). After due reflection, we decided upon the following decision rule. In order for a synthetic analysis of relevance and effectiveness at the ER level to be credible and meaningful and to allow for comparison of findings across ERs, an ER should have at least three final evaluation reports with a rating of 4 or higher on basic quality (i.e. the last variable presented in Table 2) and commissioned by an independent evaluation office. Table 3 presents the detailed criteria and decision rules for inclusion of an ER in the synthetic review of effectiveness and relevance and Annex 3 presents the detailed assessment by ER.

Table 3. Assessment framework for minimum requirements for synthetic analysis (relevance and effectiveness)

Criteria	Definitions	Scales
Evaluative Coverage	There is sufficient coverage of a programmatic area for synthesizing information on outcome and relevance	Number of Final Evaluation reports covering a given ERs.
Evaluative Quality	The evaluations have a sufficient level of quality for synthesizing information on outcome and relevance	Number of Final Evaluation reports that score 4 or 5 on the composite "quality index" (assessed in the succinct meta-evaluation).
Evaluative Independence	The evidence base stems from evaluations commissioned by a central evaluation office (whether UNESCO or not)	Number of Final Evaluation reports that were commissioned by a central evaluation office.
Minimum requirements for synthetic analysis	Composite variable, proxy of assurance for synthetic judgment on outcome and relevance summarizing information on variables: - Evaluative coverage - Evaluative quality - Evaluative independence	Dichotomous variable: yes versus no Yes if the ER is covered by at least three final evaluation reports that meet the three criteria (i.e. were commissioned by a central evaluation office and score at least a 4 on the "quality" composite index).

d. Methodological principles for the synthesis of relevance and effectiveness of UNESCO programmes

Multiple criteria should be taken into consideration when comparing areas of work in the framework of informing strategic decision-making as laid out in the [197 EX/5 Part IV], including:

- Relevance
- Capacity to deliver
- Comparative advantage/value-added
- Demonstrable contribution and tangible results
- Sustainability

To comprehensively assess UNESCO's programmatic areas of work on these dimensions, multiple sources of information are necessary: e.g., on the magnitude of funding, intervention modality, scale and outreach of UNESCO's activities, other institutions' interventions and UNESCO's 'niche position', etc.

In principle, evaluation reports have a comparative advantage in providing information on several of these dimensions. However, given the current state of evaluative coverage and quality of UNESCO's

programmatic areas of work, information on these dimensions is patchy at best. Consequently, mainly for illustrative purposes, in this synthetic review we focused on the two main dimensions of "relevance" and "effectiveness".

Relevance was assessed by looking at the following five criteria:

- The contribution of a given intervention to key global agreements, such as the MDGs and SDGs, and the EFA goals;
- The alignment with UNESCO's core priorities, as expressed in the organization's current strategic documents (C/4 and C/5);
- The alignment with national or regional priorities;
- The response to clearly identified and defined needs of the targeted beneficiaries; and
- The extent to which the intervention demonstrated UNESCO's programmatic or strategic uniqueness.

In turn, effectiveness was assessed through five main criteria as well:

- The delivery of the intervention's planned outputs;
- The achievement of outcomes;
- The demonstration of a catalytic role;
- The sustainability of effects; and
- The sustainability of partnerships.

The frameworks for assessing the relevance and effectiveness criteria are presented respectively in Tables 4 and 5.

Table 4. Assessment framework for 'relevance'

Criteria	Definition	Scale
Contribution to key global agreements (e.g., MDGs, EFA)	There is clear evidence that an intervention is well-aligned with a global agreement	Dichotomous variable: yes versus no
Response to UNESCO's core priorities	There is clear evidence that an intervention has remained relevant to UNESCO's core priorities (as expressed in current C/4 document)	Dichotomous variable: yes versus no
Alignment with national/regional initiatives	There is clear evidence that an intervention has developed synergies with related initiatives at the country/regional level	Dichotomous variable: yes versus no
Response to clearly defined needs of the targeted beneficiaries	There is clear evidence that an intervention responds to a well-targeted national or local need	Dichotomous variable: yes versus no
Uniqueness/Niche ⁸	There is clear evidence that an intervention embodies UNESCO's specificities and highlight its niche in a particular domain	Dichotomous variable: yes versus no
Relevance	Composite variable, proxy of relevance of an intervention summarizing information on variables: <ul style="list-style-type: none"> - contribution to global priorities - response to UNESCO's core mandate - addressing a specific (national/local) need 	Ordinal variable ranging from 0 to 5 AND

⁸ The concept of "niche" is used broadly to characterize areas where UNESCO has been recognized to "do something that others do not do". Apart from content analysis on statements relating to comparative advantages of UNESCO in a particular area of work, we also counted them.

	<ul style="list-style-type: none"> - synergy with related initiatives at the country/regional level - uniqueness/niche 	Ratio of reports scoring 3 or more on 5 dimensions
--	--	--

Table 5. Assessment framework for 'effectiveness'

Criteria	Definitions	Scales
Output delivery	There is clear evidence that the intervention delivered most planned outputs	Dichotomous variable: yes versus no There is often a lack of clarity on the exact wording for output. Terms like outputs, deliverables, results may be used to refer to outputs. Here we focus on products and services delivered or resulting from an intervention.
Outcome achievement	There is clear evidence that the intervention is likely to achieve intended (or unintended) positive outcomes	Dichotomous variable: yes versus no Yes if: Positive outcome on one (or more) main intended outcomes and no indication of negative outcome OR positive outcomes clearly outweigh minor negative outcomes No: otherwise
Catalytic role and effect	There is clear evidence that an intervention is likely to have a catalytic effect	Dichotomous variable: yes versus no A catalytic effect can be simply defined as an instance where a relatively small factor provokes substantial changes (e.g., instances of nudging effect, scaling up, or replication).
Sustainability of effects	There is clear evidence that the intervention will lead to long-term sustained change	Dichotomous variable: yes versus no Yes if: clear indication of sustainability of effects (including supportive evidence that underlying assumptions are met) No: otherwise (sustainability unlikely to be achieved as assumptions are not met or clear indications that effects will not be sustained).
Sustainability of partnerships	There is clear evidence that the intervention is supported by partners beyond the timeframe of the UNESCO intervention	Dichotomous variable: yes versus no
Effectiveness	Composite variable, proxy of effectiveness, summarizing information on variables: <ul style="list-style-type: none"> - output delivery - outcome achievement - catalytic effect - sustainability of effects - sustainability of partners 	Ordinal variable ranging from 0 to 5 AND Ratio of reports scoring 3 or more on 5 dimensions

MaxQDA software was used to support the tasks of coding, extracting, and synthesizing evidence stemming from evaluation reports. Using the software also enhanced the transparency and traceability of the synthetic steps.

When performing data extraction and rating, applying the two matrices presented above, the following assumptions and decisions were made:

- Results can be traced back to UNESCO's contribution. This is more straightforward in the case of evaluations of interventions where UNESCO is the sole, or primary actor. However, in many evaluations, UNESCO is only one of the contributing (funding/implementing) actors. While dealing with this issue could be resolved by weighing the outcome achievement by a simple metrics of UNESCO's contribution, on a three-point scale (e.g., main contributor, one of several key contributors, small contributor), this might not be desirable as encouraging partnerships and joint interventions is a key objective of the institution;
- Both intended and unintended outcomes that are mentioned in the report are taken into account; and
- Quantity and scale of output delivery and outcome achievement (e.g. the scale and diversity in outputs, the number and magnitude of intended outcomes of an intervention) are not (directly) taken into account.

The minimum requirements for synthetic analysis were met only for a small subset of UNESCO ERs. Consequently, a comprehensive cross-programmatic comparison of relevance and effectiveness was not possible. Therefore, a synthetic analysis of relevance and effectiveness was performed on one Expected Result per sector for which the minimum requirements were met. For these five ERs, we extracted information from the evaluation reports covering each ER, focusing on relevance and effectiveness. The sample of evaluation reports for this part of the analysis amounted to 49 final evaluation reports.

The assessment matrices for relevance and effectiveness were applied to all evaluation reports pertaining to one of the five ERs retained for illustrative purposes. The extracted information was then synthesized qualitatively and formed the basis for rating the ERs on the relevance and effectiveness criteria. Finally, at the level of the composite indicators for effectiveness and relevance, the share of reports with a rating of at least three out of five was computed. This analysis, which was conducted for illustrative purposes, should ideally be scaled up to cover the entire UNESCO portfolio if certain conditions are met (see the report's recommendations).

e. Methodological principles for the synthesis of cross-cutting challenges and enabling factors

The reports reviewed in the synthesis of relevance and effectiveness were also systematically screened for information on enabling and challenging factors that affect UNESCO's programme delivery and processes of change. In addition to these reports, all corporate evaluation reports that dealt with cross-cutting issues were included in this last part of the analysis, resulting in a sample of 68 reports.

A coding system for enabling and constraining factors was developed and systematically applied to the sample of reports. Finally, results were synthesized to the level of the Organization.⁹

3. Limitations

A number of methodological caveats inherent to this type of synthesis exercise should be kept in mind. Broadly these caveats can be divided into three categories:

1. the (lack of) representativeness of the sample of evaluation reports reviewed;
2. the (difficulties of) mapping evaluation reports onto programmatic areas, and;

⁹ Despite certain biases in the underlying sample of reports, it was hypothesized that due to the "systemic" nature of many of the constraints and enabling factors, patterns of prevalence could be generalized to the level of the UNESCO system.

3. the (limited) depth of the analysis.

Caveats related to sample biases

- UNESCO does not currently have a platform and a procedure to systematically collect decentralized evaluation reports and to comprehensively account for the entire universe of evaluation reports conducted within the UNESCO system. This means that gaps in the coverage analysis could be due to one (or both) of the following reasons: (i) because the reports exist but were not obtained by IOS; or (ii) because the reports do not exist.
- The most common type of decentralized evaluation is at the project level, mostly donor-funded. It is important to highlight that there have been guidelines on the evaluation of extrabudgetary activities for a long time within UNESCO. These guidelines have recently been modified and are part of the overarching evaluation policy, which reiterates that "all UNESCO extrabudgetary activities are subject to evaluation [and] that an external evaluation is mandatory for all extrabudgetary projects with an allocation greater than \$1.5 million." However, the guidelines are not yet fully implemented, as shown in the 2013 Diagnostic Study.
- In theory, it is possible to determine the extent to which the guidelines were implemented. But within the current system, this is a very tedious and time-consuming task. It is, thus, difficult to establish the extent to which the sample of external evaluations is sufficiently comprehensive and representative of the total population of external evaluations conducted in the period covered by the present exercise. The only conclusion that we can draw is that the sample is reasonably diverse in terms of coverage of sectors and the time period of the study. However, there are clear sample biases that should not be overlooked. To cite only one example, evaluations of MDG-F activities are overly represented in the total sample.
- While decentralized evaluations usually relate to extrabudgetary projects, corporate evaluations focus on central services, institutional entities (e.g. Cat I Institutes), themes or programmatic areas of work that are supported by the Regular Programme Budget as well as extrabudgetary resources. In contrast to corporate evaluations, decentralized evaluations are not bound to a predetermined plan to cover particular areas of (extrabudgetary work).

Caveats related to mapping evaluations onto ERs

Mapping the evaluation reports onto ERs was a difficult exercise that required a significant amount of judgment, interpretation, and extrapolation. A number of reasons account for this:

- ERs are formulated on the basis of political consensus-building processes, resulting in a high number of intended results that often lack the precision and accuracy.
- UNESCO currently not have a consistent definition of a programme as a unit of intervention (and a unit of analysis). Correspondingly, there is no standardized programme information cycle¹⁰ to support reporting and decision-making. Both are important building blocks for inter-programme comparisons of merit and worth.
- In addition, ERs are not consistently formulated at the outcome level. As demonstrated by the 2014 Formative Evaluation on UNESCO Results-Reporting, the causal language regarding results in the formulation of ERs is not consistent across ERs. As noted in previous studies, the concrete formulation of an "ER" in practice differs. Sometimes it is formulated as an outcome or even impact, but sometimes it also takes the form of an output, or even a description of activities.¹¹
- A large number of evaluation reports cover extrabudgetary activities. There is no systematic effort to articulate intervention logics at the level of programmatic areas of work and to link

¹⁰ Similar to the project information cycle: ex ante, interim and ex post data collection and analysis activities, including external evaluations.

¹¹ However, recent improvements in the formulation of ERs in the 37 C/5 should be noted.

extrabudgetary activities to these intervention logics. Consequently, it is difficult to map evaluation reports of extrabudgetary intervention to ERs with a certain degree of precision and confidence.

- UNESCO does not have a consistent way of accounting for results achieved through intersectoral collaboration or activities within sectors that are intersectoral in nature. There is no intervention logic or results framework that would clearly lay out how intersectoral work is meant to contribute to single or multiple ERs. This weakness translates in important challenges to assess intersectoral work in evaluations and in accurately mapping evaluation reports onto ERs.
- A large number of evaluation reports cover joint programmes that were implemented in the framework of the MDG-F and are at times only loosely connected to UNESCO's strategies and programmes.
- A number of evaluation reports cover programmes that were implemented prior to the current C/5. While the MLA level has remained quite stable over time, there has been a marked decrease in the number of ERs between the 36 C/5 and the 37 C/5.

Caveats related to the level of analytical depth

- Given the large sample of evaluation reports to be reviewed, a thorough analysis of quality was not possible and the meta-evaluation relied on a range of proxy measures to assess quality and independence, some of which are rather superficial.
- Measuring effectiveness (which relates to the outcomes of an intervention) is particularly challenging due to, among other things: missing information, the timing of effects (which can take several years to come about), the timing of the evaluation (which takes place often just after completion of an intervention), and the evaluation methods used. Consequently, references to outcome achievement in evaluation reports generally concern assessments of the “likelihood” that outcomes will be achieved in the longer term.
- There is an imbalance in evaluation reports towards providing more thorough and in-depth analyses of factors that were found to hinder implementation processes and processes of change, rather than enabling factors. This imbalance is unavoidably reflected in this synthesis, with more examples of challenges than enabling factors.

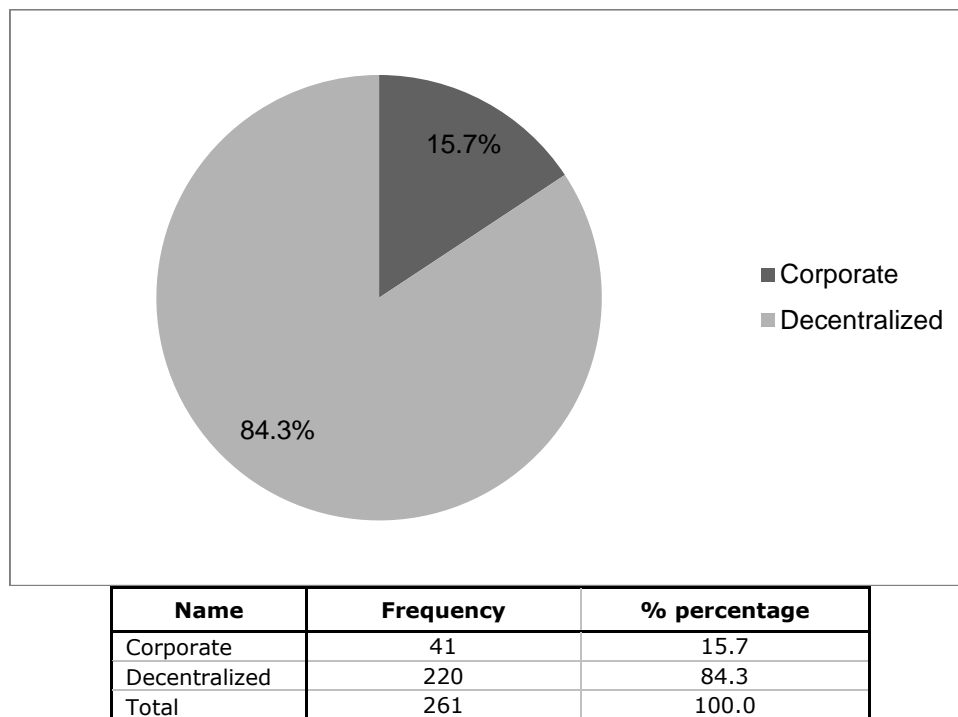
CHAPTER 2: COVERAGE ANALYSIS

In this chapter, we present a diagnostic overview of the landscape of evaluations in the UNESCO system. First, we provide an overview of the sample of evaluations at the level of the organization. Subsequently, we present a basic description of our sample, looking among other things at the sectoral distribution of reports. The core part of the analysis is our mapping of evaluation reports onto ERs that will serve as a basis for the remaining of the study.

1| COVERAGE AT THE UNESCO-WIDE LEVEL

Our data collection and analysis covers the period January 2008 to September 2015 for corporate evaluation reports, and January 2009 to September 2015 for decentralized evaluation reports. We collected a total of 261 reports. As shown in Figure 1, the majority of the collected sample is constituted of decentralized evaluations (84.3%) and a smaller proportion (15.7%) of corporate evaluations (commissioned and/or conducted by IOS).

Figure 1. Total number of evaluation reports by type (n = 261)



As illustrated in Figure 2, the majority (70.9%) of reports in the sample are final evaluation reports.¹² Less than a third (27.6%) of reports are mid-term evaluation reports.

¹² To keep things simple, corporate evaluations are considered as final evaluations (even if they concern ongoing interventions or portfolios of interventions).

Figure 2. Total number of evaluation reports by timing (n = 261)

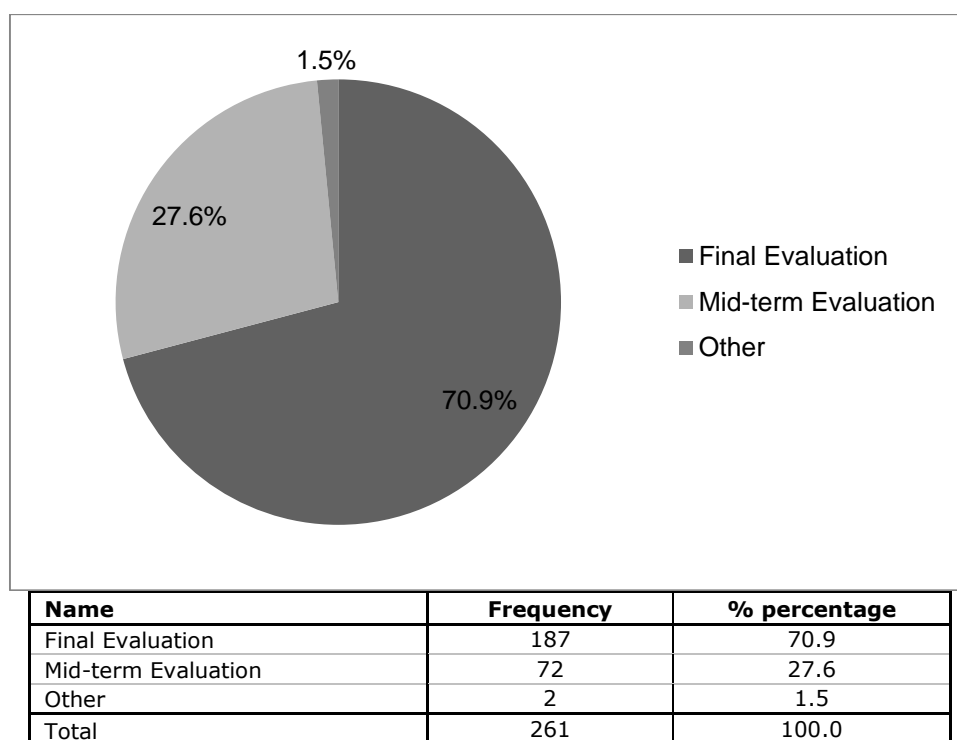
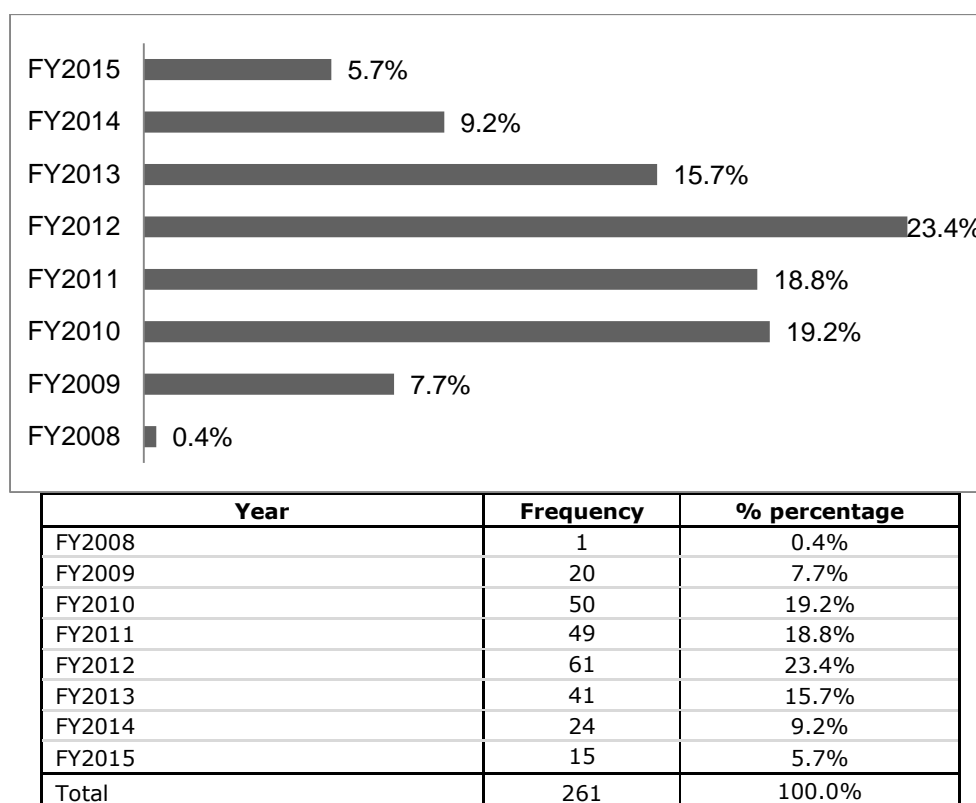


Figure 3 illustrates the distribution of reports by year of publication.¹³ The figure shows that the sample has a decent coverage of the years 2009 to 2013 while there are substantially less reports for 2014 and 2015. In addition to the list of caveats laid out in Chapter 1 with regards to sample bias, the main reasons for the lower number of reports after 2013 are threefold:

- There has been an elaborate effort to collect evaluation reports for the 2013 Diagnostic Study.
- There has been a surge of evaluation reports in 2012 linked to the closure of many of the interventions in the framework of the Culture and Development window of the MDG-F.
- The data collection phase for this exercise was concluded in September 2015, partially explaining the smaller number of evaluations for this fiscal year compared to other years.

¹³ As explained above, the sample only covers corporate evaluations for the year 2008 (due to the decision to incorporate all SPO evaluations), which explains why only one report is included in this year.

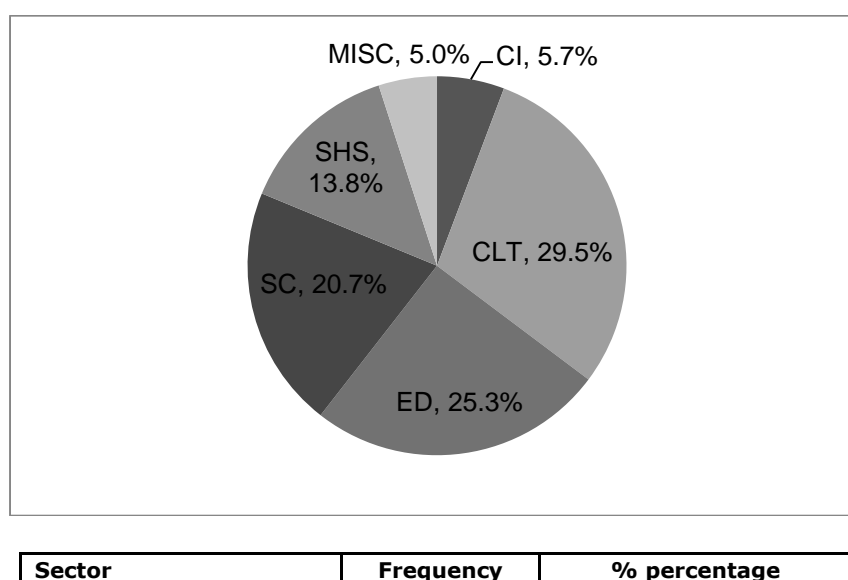
Figure 3. Total number of evaluation reports by year of publication (n = 261)



2| COMPARISON ACROSS SECTORS

Figure 4 presents the sectoral distribution of reports. The majority of reports cover the Culture (CLT) sector (29.5%) and the Education (ED) sector (25.3%). The Natural Sciences (SC) sector comes third with a fifth of the evaluation reports (20.7%). The Social and Human Sciences (SHS), and the Communication (CI) sectors have a minor share of the evaluation reports with respectively 13.8% and 5.7% of reports. Finally, roughly 5% of reports cover transversal themes or corporate services (MISC).

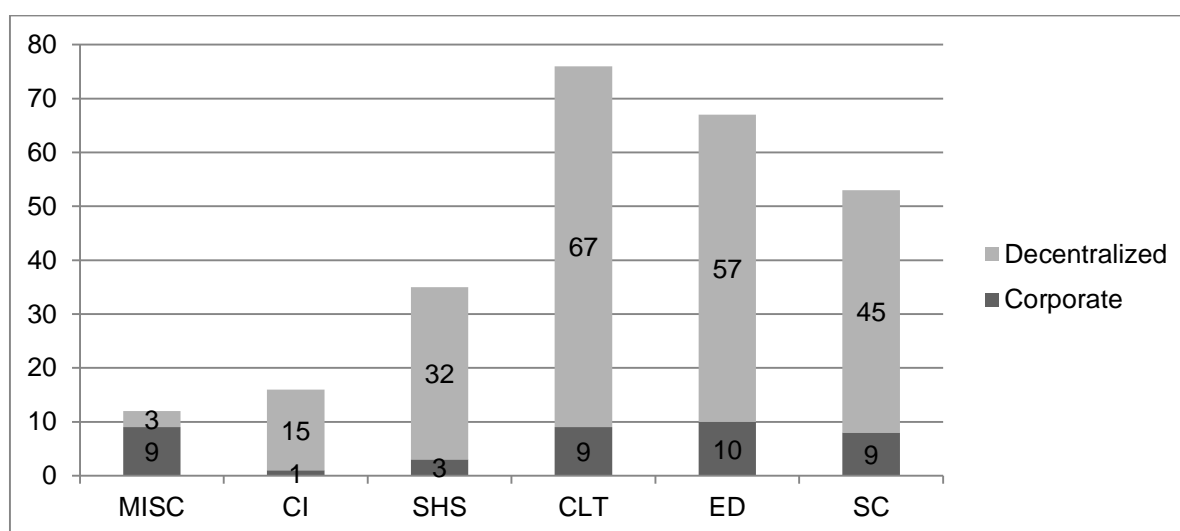
Figure 4. Total number of evaluation reports by sector (n = 261)



CI	15	5.7%
CLT	77	29.5%
ED	66	25.3%
SC	54	20.7%
SHS	36	13.8%
MISC	13	5.0%
Total	261	100.0%

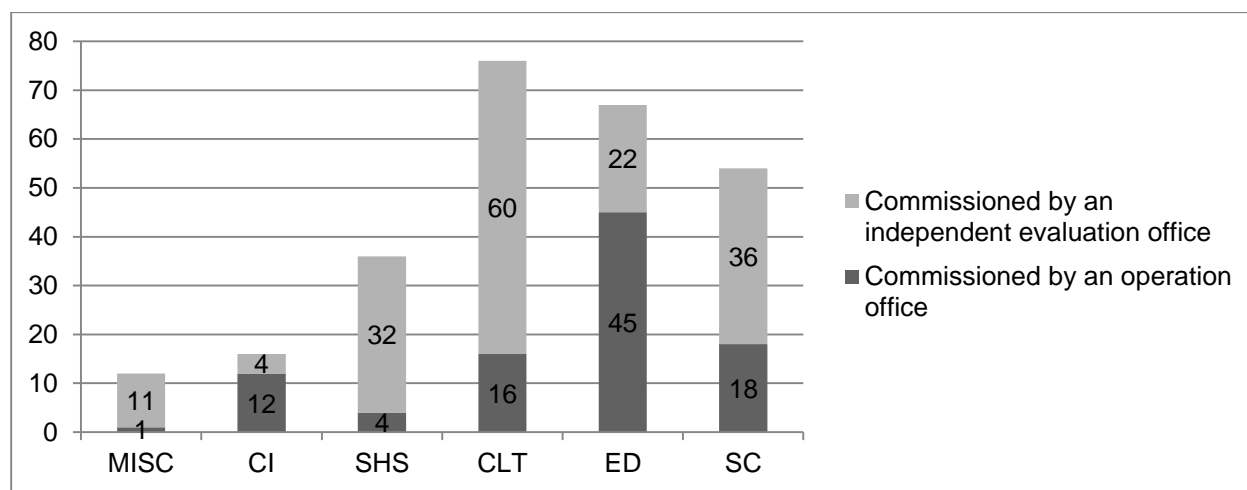
In Figure 5 we show the distribution of corporate versus decentralized evaluation reports disaggregated by sector. The figure clearly shows that while the CLT, ED, and SC sectors are well represented in corporate evaluation reports, the CI and SHS sectors have been subject to a lower coverage by IOS, with only one dedicated corporate evaluation in the CI sector in the past 7 years.

Figure 5. Total number of evaluation reports by type and sector (n = 261)



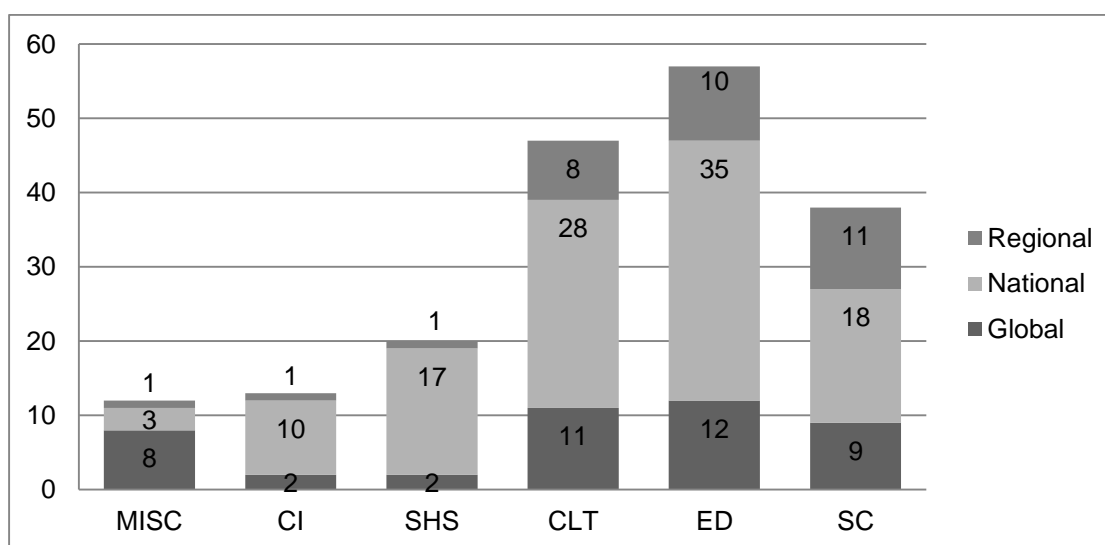
As a rough proxy for the level of independence of the evaluation, we consider whether the report was commissioned by the unit in charge of operations or by an independent evaluation office (e.g., IOS, the MDG-F central evaluation function, or a donor evaluation office). Figure 6 illustrates the distribution of evaluation reports by sector for this variable. There are substantial differences across sectors. While the CI and ED sectors have most of their evaluations commissioned by operations, the CLT, SC, and SHS sectors have most of their evaluation commissioned by a central evaluation office (within or outside of UNESCO).

Figure 6. Total number of evaluation reports by type of commissioner (n = 261)



As noted in the 2013 Diagnostic Study, the level of implementation of UNESCO activities can be a useful proxy for distinguishing between projects that are mainly managed by Headquarters versus activities managed mainly by Field Offices. Figure 7 shows that the majority of reports correspond to national (or sub-national) level evaluands. A minority of projects relate to interventions of a regional (multiple country) or global scope.

Figure 7. Total number of evaluation reports by level of implementation (n = 261)

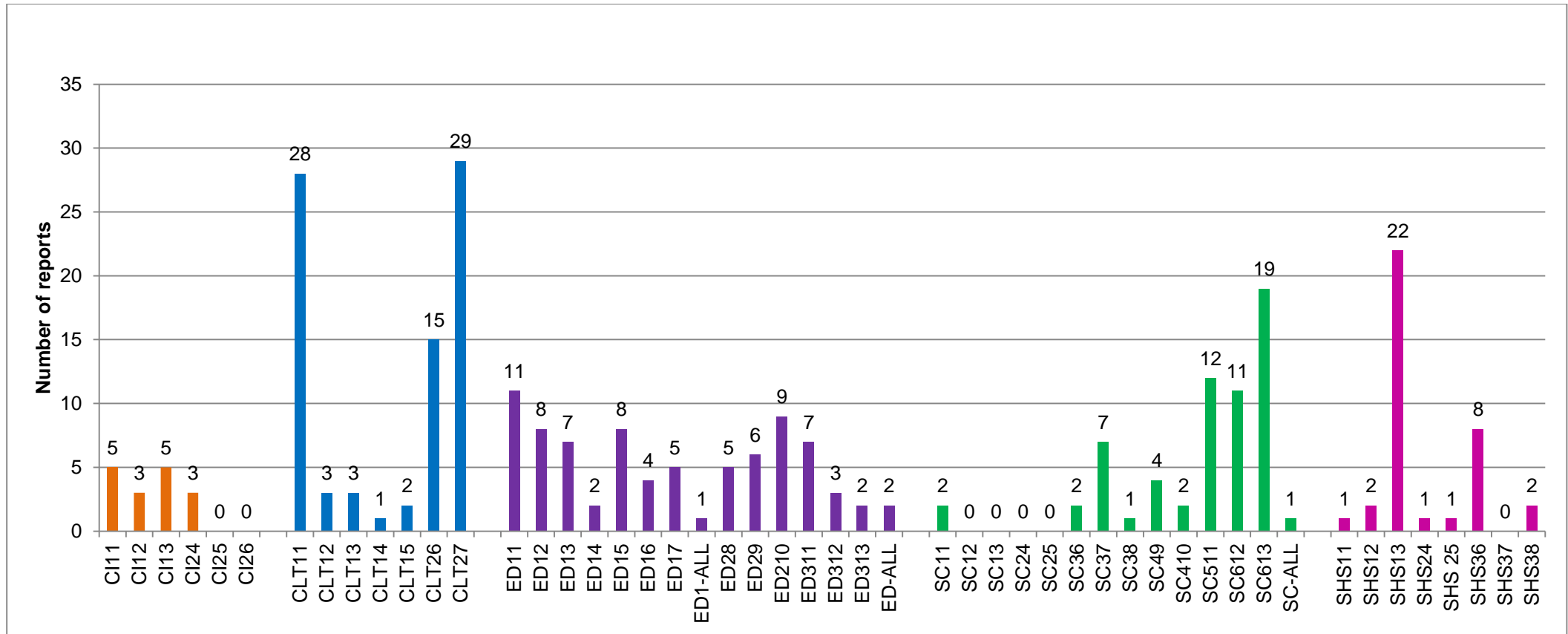


3 | COVERAGE WITHIN SECTORS

As a preliminary condition for a synthetic analysis of effectiveness and relevance of UNESCO programmatic areas, we mapped evaluation reports onto the 37C/5 ERs for each sector. Given that some evaluations have a broad scope and cover multiple ERs, these reports were mapped onto each corresponding Expected Result. As shown in Table 6, the total number of reports thus exceeds the actual sample of reports (n = 261) because a single evaluation report can be mapped onto several ERs, thus counted multiple times. In Table 6 we also distinguish between reports that are within a single Expected Result and those that cover multiple ERs.

As illustrated in Figure 8, in each sector, there are both ERs with substantial evaluation coverage, and ERs with minimal or no evaluation coverage. The ERs for which there is the largest number of evaluation studies can be found in the CLT sector, with two ERs with more than 25 evaluation reports. The CLT and ED sectors both achieve at least a minimal evaluation coverage of all of their ERs (at least one evaluation report), whereas the CI, SC, and SHS sectors all have at least one ER with no evaluation coverage.

Figure 8. Total number of evaluation reports by Expected Results (n = 261*)



Notes:

* A single evaluation report can be mapped onto several ERs, and thus be counted multiple times. As a result, the total in this figure (287) exceeds the actual sample size (n = 261).

Table 6. Total number of evaluation reports by Expected Results (n = 261*)

<i>ERs</i>	<i>ER Code</i>	<i>Within ER</i>	<i>Multiple ER</i>	<i>Total by ER</i>
Cross-cutting themes and corporate services	MISC		12	12
Communication				
Freedom of expression, press freedom and journalistic safety	CI11	4	1	5
Community and pluralistic media and media literacy among youth	CI12	2	1	3
IPDC	CI13	5		5
Open Solutions and ICT	CI24	3		3
Memory of the world programme	CI25			0
World Summit on the Information Society (WSIS) and the Information for All Programme (IFAP)	CI26			0
Culture				
Tangible heritage: 1972 convention	CLT11	19	9	28
Illicit transfer of cultural property: 1970 convention	CLT12	2	1	3
Protection of cultural property in the event of conflict: 1954 convention	CLT13	3		3
Protection of underwater cultural goods: 2001 convention	CLT14	1		1
History and memory	CLT15	1	1	2
Intangible cultural heritage, indigenous and endangered languages: 2003 Conventions	CLT26	11	4	15
Diversity of cultural expressions: 2005 conventions	CLT27	26	3	29
Education				
Education Policy and planning	ED11	8	3	11
Literacy	ED12	8		8
TVET	ED13	6	1	7
Higher Education	ED14	1	1	2
Teacher policies and strategies	ED15	6	2	8
Promotion, monitoring of competency-based learning	ED16	3	1	4
ICT and Education	ED17	4	1	5
All MLA 1 on Supporting Member States to develop education systems to foster high quality and inclusive lifelong learning for all	ED1-ALL		1	1
Peace and human rights education	ED28	5		5
Education for Sustainable development	ED29	5	1	6
Health Education	ED210	9		9
Foresight and research for EFA	ED311	0	7	7
Monitoring of EFA	ED312	2	1	3
Cooperation and partnerships for EFA	ED313	0	2	2
All of the Education Sector	ED-ALL		2	2

(Continued)

<i>ERs</i>	<i>ER Code</i>	<i>Within ER</i>	<i>Multiple ER</i>	<i>Total by ER</i>
Natural Science				
STI policies	SC11	1	1	2
Sciences-Policy interface and Sustainability science	SC12			0
Vulnerable groups, SIDS and indigenous peoples	SC13			0
Capacity building in research and education in natural sciences	SC24			0
Engineering research	SC25			0
Ocean and coastal research	SC36	1	1	2
Ocean-related hazards and climate change (e.g., tsunamis)	SC37	6	1	7
Institutional capacity for management of ocean and coastal resources	SC38	1		1
Global cooperation in ecological and geological sciences	SC49	3	1	4
Early warning of natural hazards (e.g., earthquake)	SC410		2	2
Biosphere reserves	SC511	12	0	12
Water security challenges (e.g., drought, flood)	SC612	10	1	11
Capacity for water management	SC613	16	3	19
All of the Natural Science Sector	SC-ALL	1		1
Social & Human Science				
Social transformation	SHS11	1		1
Intersectoral work for social inclusion	SHS12	2		2
Peace and security and intercultural dialogue	SHS13	22		22
Bioethics	SHS24		1	1
Ethics	SHS 25		1	1
Youth	SHS36	8		8
Sports and physical education	SHS37			0
Human rights based approach	SHS38	2		2
	Total *	220	67	287

Notes:

* A single evaluation report can be mapped onto several ERs, and thus be counted multiple times. As a results, the total in this table (287) exceeds the actual sample size (n = 261).

4| KEY FINDINGS

Notwithstanding the caveats in terms of sample bias laid out in Chapter 1, there are two clear findings emerging from this coverage analysis.

Finding 1: Evaluative coverage is uneven across sectors: while some UNESCO sectors demonstrate a rather good evaluation coverage (Culture, Education and Natural Sciences), other sectors have a very low evaluation coverage (Social and Human Sciences, Communication and Information).

Finding 2: Within sectors, evaluation coverage is very uneven across ERs: while some ERs are particularly well-covered, others are essentially “evaluation-free” with no evaluations at the level of ERs or no “within ERs” evaluation. This is the case for all UNESCO sectors.

CHAPTER 3: SUCCINCT META-EVALUATION

This chapter is dedicated to assessing a number of variables related to the quality of evaluation reports. The sample for this review comprises all of the final evaluation reports (excluding mid-term evaluation reports and including both decentralized and corporate reports). We focus on the content of the report, which covers a number of dimensions that can be considered as proxies for quality (see methodology section): implementation, outputs, effects, lessons learned, clarity of the causal chain, and presence of a logical framework (logframe) or logic model. To be consistent with the Diagnostic Study carried out in 2013, a composite variable of “quality” was developed to synthesize the information on the previous five variables. This proxy measure for quality is analyzed over time and across sectors. In addition, we created another composite measure which determines the “minimum requirements for synthetic review” assessed at the level of each programmatic area. This composite measure combines the composite measure of quality, a proxy measure for the independence of evaluative judgment, and a measure of coverage (described in the previous chapter).

1 | BASIC QUALITY OF EVALUATIONS

The measures discussed in this section are all dichotomous variables (except for the composite variable "quality" which is measured on an ordinal scale). In order to ensure continuity with previous assessments, we used the same straightforward decision rules as applied in the 2013 Diagnostic Study. Any information on implementation, outputs, effects, etc. found in a report, no matter how much, was sufficient to obtain a "yes" on the variable. As a result, it is not surprising that with few exceptions, all the reports got a "yes" on implementation and outputs, which constitute the core of the analysis in final evaluation reports.

Information on implementation and outputs

Notwithstanding the above, the depth and coherence of the discussion of outputs is often unsatisfactory. While we did not review the entire sample in sufficient detail to be able to systematically assess the quality of the information on each dimension, we did so for a significant sub-sample as preparatory work for deciding the sampling strategy. We can thus conclude with confidence that overall the discussion on the quality and quantity of outputs (deliverables) remains rather weak.

Information on effects

While implementation and outputs information is included in all reports, this is not the case for information on effects (i.e. the intended or unintended changes that result directly or indirectly from a given intervention). That being said, the large majority of final evaluation reports contain some information on effects. As shown in Figure 9 only 17% of the reports in the sample did not include any discussion on effects. Again, a word of caution is in order. While the large majority of reports include a cursory discussion of effects, a large proportion of reports discusses effects superficially with key aspects missing from the reports such as:

- The full range of direct or indirect effects is hardly discussed.
- The effects are described in broad terms, lacking specific qualitative or quantitative evidence on the nature and scale of effects (what changed, for whom, to what extent).
- A discussion of contribution or attribution is often missing, with little or no attention being paid to alternative explanations for observed changes.

Figure 9. Total number of final evaluation reports by effects (n = 187)

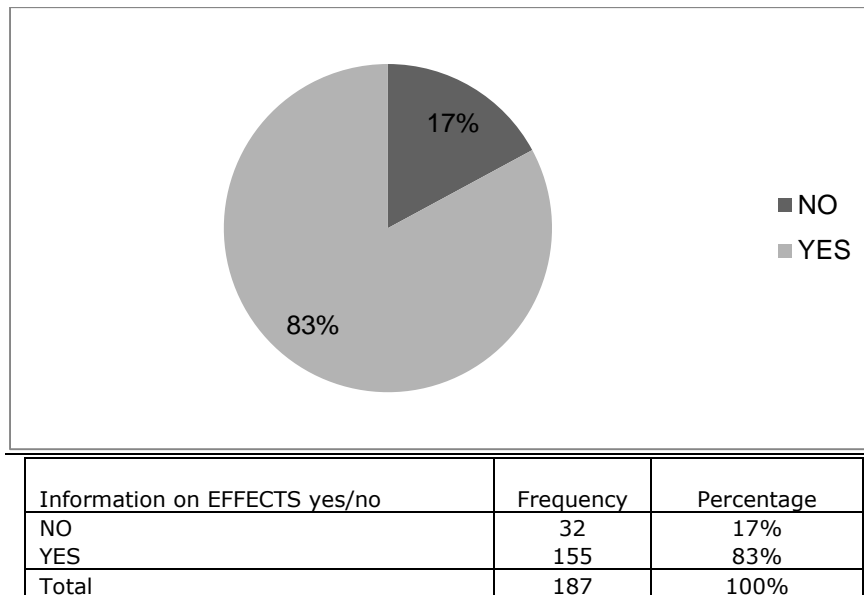
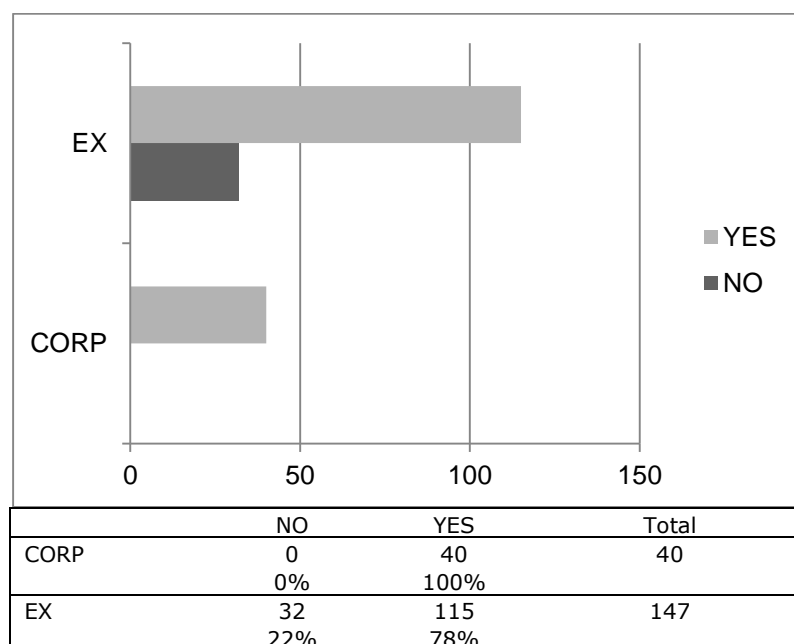


Figure 10 shows a clear difference between decentralized evaluations and corporate evaluations in terms of the extent to which they cover information on effects. While all the corporate evaluations systematically include at least some discussion of effects, only 78% of the decentralized reports do so. The difference between the two types of reports is not surprising given the absence of systematic quality control of decentralized evaluation reports.

Figure 10. Total number of final evaluation reports by effects and by type (n = 187).



Information on lessons learned

Another important variable of quality is whether a report includes some reflection on lessons learned from the intervention, whether it is a discussion of challenges, opportunities, or considerations that should be taken into account in future operations. The variable was marked as "yes" if there was a

dedicated chapter, section, or paragraph within the report that included some of these elements. As displayed in Figure 11 the majority of reports (87%) do.

Figure 11. Total number of final evaluation reports by lessons learned (n = 187)

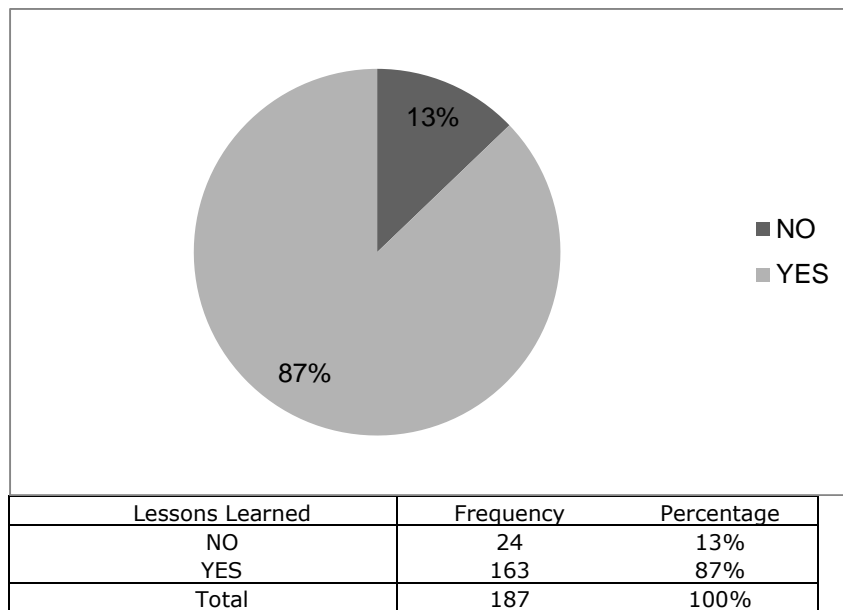
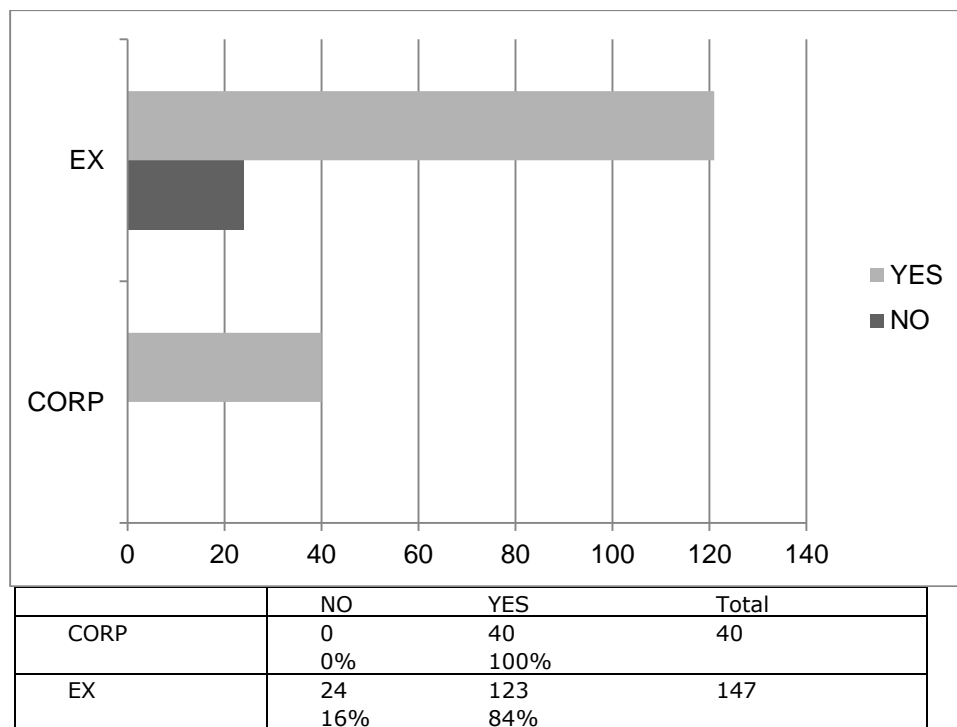


Figure 12. Total number of final evaluation reports by type and by lessons learned (n = 187)



As expected, the data show a similar pattern for the inclusion of information on lessons learned as it did for information on effects. All corporate evaluations include at least a cursory discussions of lessons learned in comparison to 84% of decentralized evaluations (Figure 12).

Clarity of the causal chain

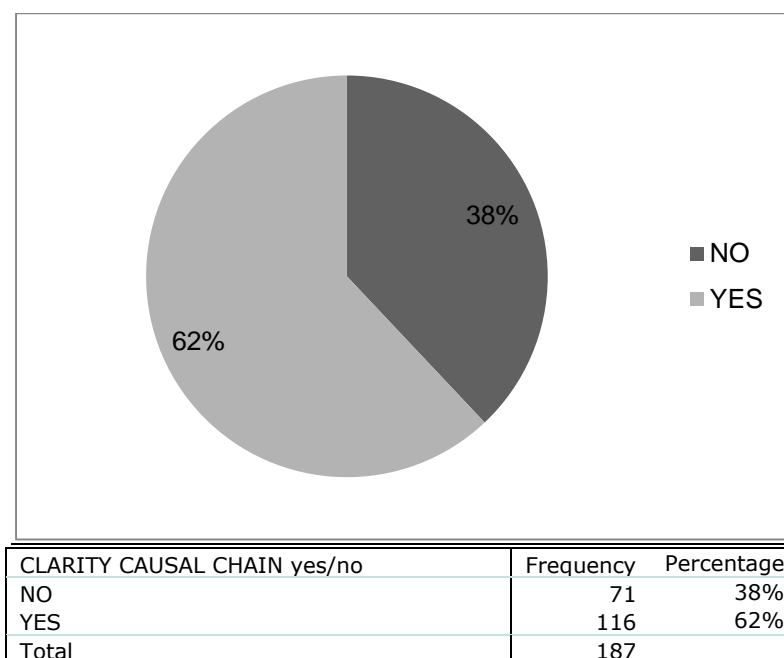
The variable "clarity of causal chain" captures whether or not a report uses causal language consistently. The decision rules for this variable were applied in two steps (consistent with the methods used in the 2013 Diagnostic Study):

1. We identified whether information on the three elements of implementation, outputs and effects was included in the report.
2. We assessed whether there was a clear distinction between these three elements (no matter the terms used to qualify effects) and a coherent use of terminology throughout the report (systematic use of the same term to mean the same thing).

If both elements were present then we rated the project as "yes" on this variable.

We find that a substantial proportion (38%) of the overall sample of reports has an issue with clear and consistent terminology on the causal chain, as displayed in Figures 13 and 14.

Figure 13. Total number of final evaluation reports by clarity of causal chain (n = 187)



Looking at decentralized evaluation reports only, we see that 52% of the reports use clear causal language while 48% exhibit some inconsistency in the use of causal language and/or do not cover all of the three elements (implementation, outputs and effects) in the report.

Further evidence on clarity of the causal chain is provided by the variable logframe (presence of logical framework or logic model in the report). Of all final evaluation reports, only 40% contain some type of logical framework or logic model (Figure 15).

Figure 14. Total number of final evaluation reports by clarity of causal chain and type (n = 187)

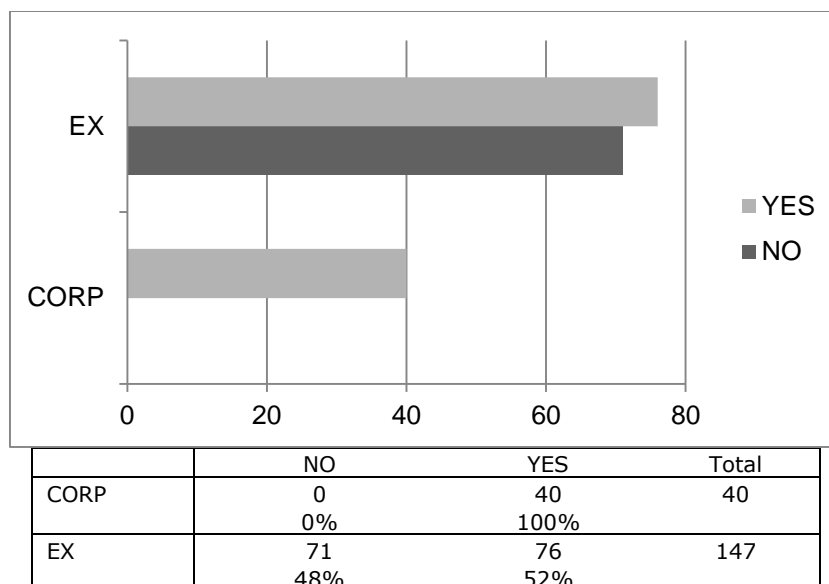
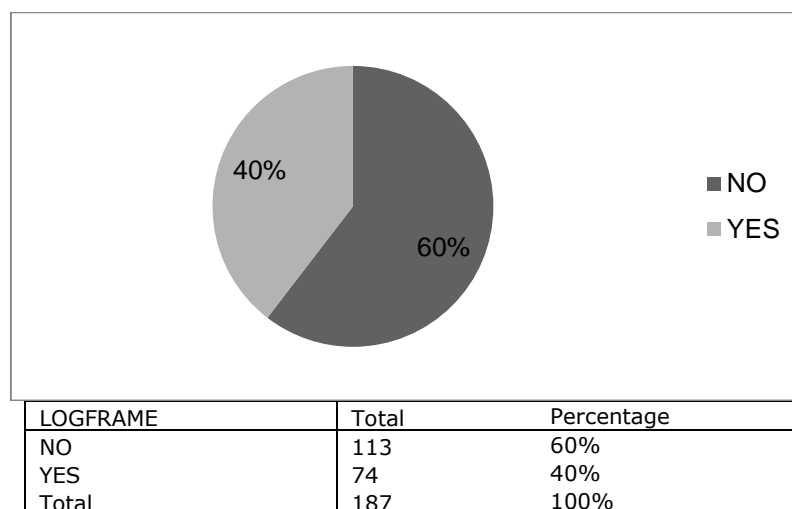


Figure 15. Total number of final evaluation reports by logframe (n = 187)



Composite variable: proxy for quality

The composite variable "quality" summarizes the information of the previous five variables into a single score, on a scale from 0 to 5. Figure 16 shows that there is a clear difference between the distribution of quality scores by type of evaluation. While all corporate evaluations receive a score of 5 out of 5 on the composite quality indicator, the decentralized evaluations are more likely to receive a 4 out of 5 on the composite variable (with a mean of 4.096). The way we defined the variables and applied the decision rules is consistent with the 2013 Diagnostic study, and thus subject to the same caveat: the composite variable captures *the minimum requirements for quality* rather than a *guarantee* of quality. In order to develop a more nuanced measure of quality, a detailed content analysis would have to be performed on each report, which is outside the scope of this exercise.

Figure 16. Total number of final evaluation reports by type and by quality points (n = 187)¹⁴

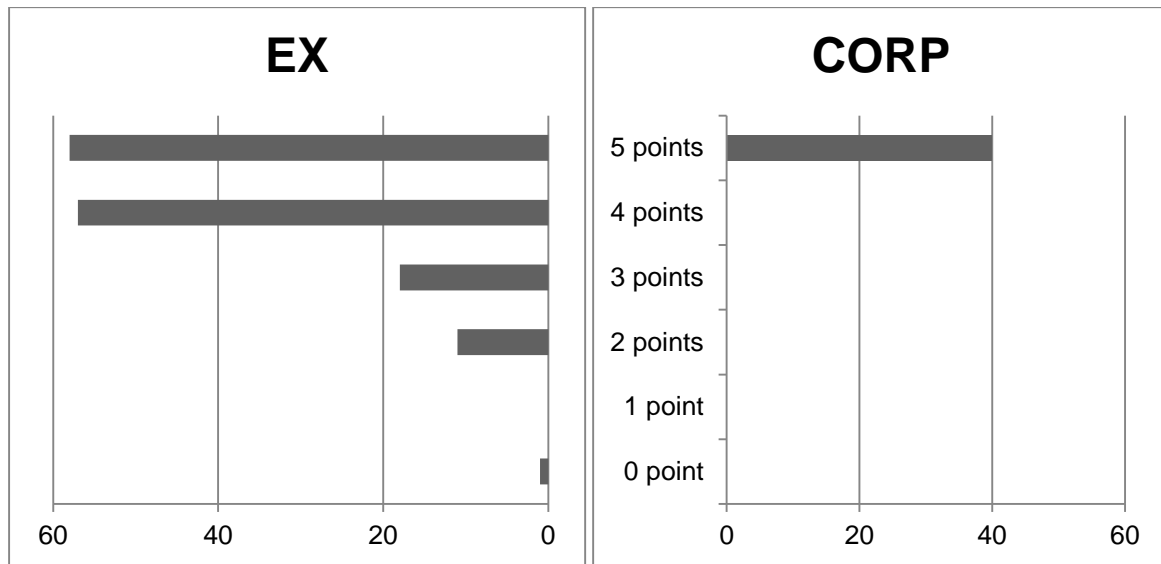
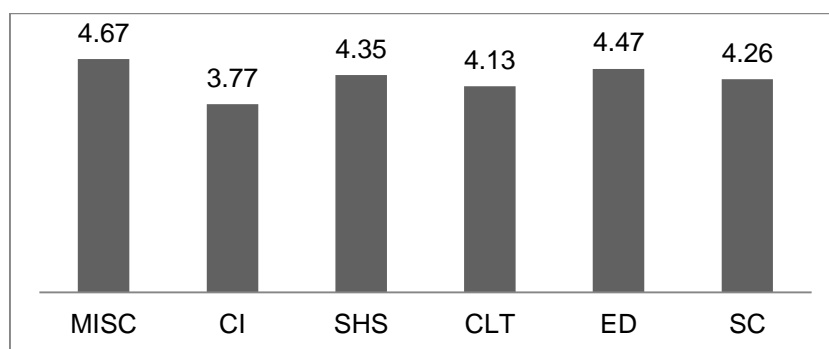


Figure 17 presents the average quality of final evaluation reports for the different UNESCO sectors. Apart from a lower average in the CI sector (which has to be interpreted cautiously due to a significantly smaller sample size than for the other sectors), there is no clear pattern at the level of the sector emerging from this comparison (see panel a). However, when removing corporate evaluations from the sample and considering only decentralized final evaluation reports, a pattern emerges. As shown in panel b of Figure 17, the average quality of reports is the highest in the ED, SHS, and CLT sectors. The high average quality of SHS reports in this sample can be explained both by the smaller sample size for this sector and by the large proportion of reports commissioned in the framework of the MDG-F programme.

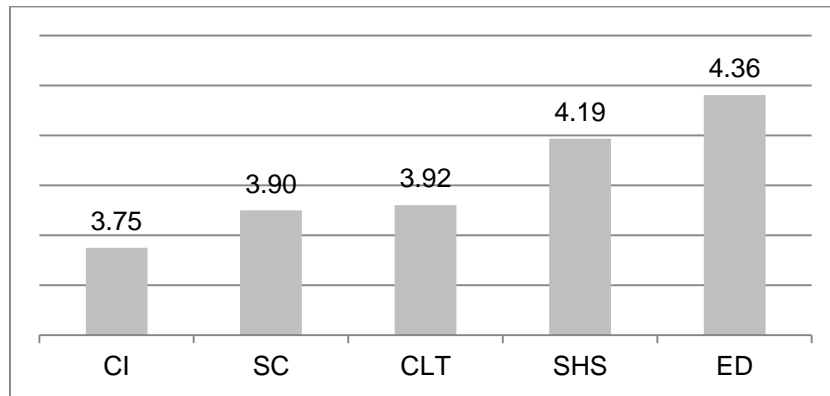
Figure 17. Average quality points of final evaluation reports by sector

a) Total number of final evaluation reports by average quality points and by sector (n = 187)



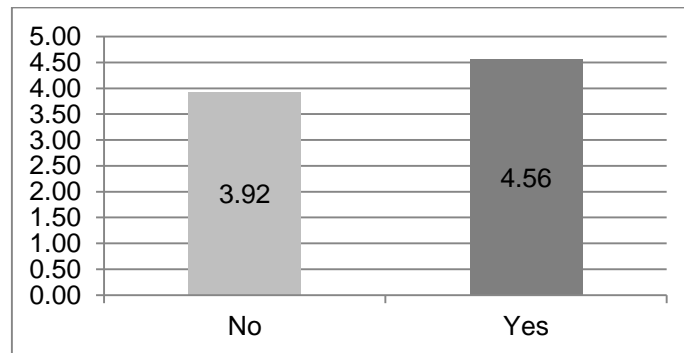
¹⁴ One of the decentralized evaluation reports was titled "final evaluation" but did not apply any of the conventional evaluation criteria. The short report was written as a technical expert assessment on the technical soundness of a project.

b) Total number of decentralized final evaluation reports by average quality points and by sector (n = 146)



As displayed in Figure 18 there is a substantial difference in the average quality of final evaluation reports that were commissioned by a central evaluation office and those that were not. While the former show an average 4.56 out of 5, the latter show an average of 3.92.

Figure 18. Total number of final evaluation reports by average quality points and by type of commissioner: "independent"= yes (n = 187)



2| BASIC QUALITY OVER TIME

It is particularly interesting to analyze whether there has been a change in quality between the sample of the 2013 Diagnostic Study and more recent reports. Given that the 2013 Diagnostic Study only reviewed decentralized evaluation reports, in this analysis we disregard the 40 corporate evaluation reports and focus on the 147 decentralized final evaluation reports only.

Figure 19 shows the distribution of quality points for the 2013 sample and Figure 20 shows the quality distribution for the sample of more recent decentralized final evaluation reports.

Figure 19. Distribution of average quality points for the sample of final evaluation reports included in the 2013 Diagnostic study (n = 78)

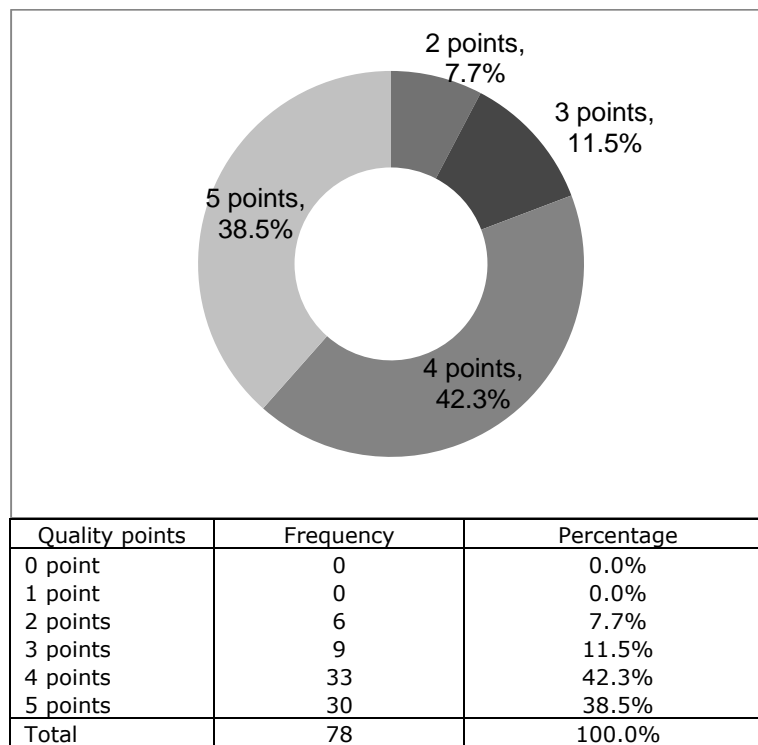
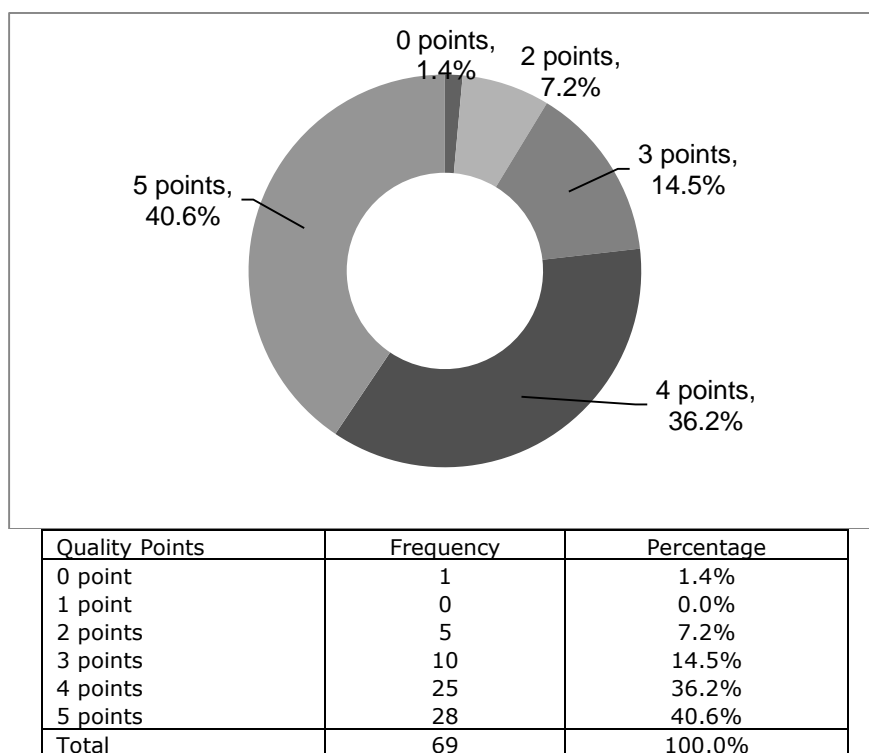
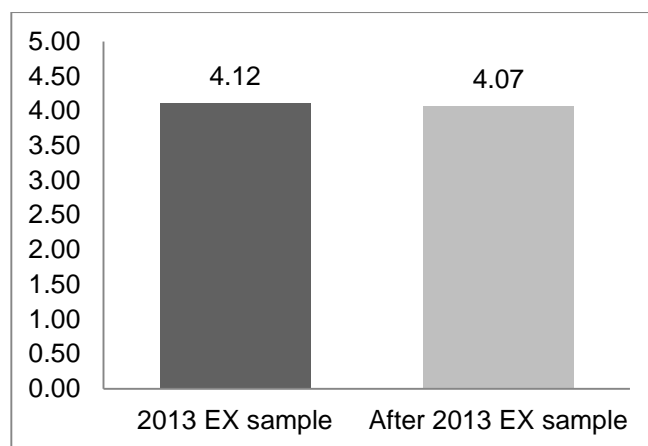


Figure 20. Distribution of average quality points for the sample of recent final evaluation reports (after 2013; n = 69)



As shown in Figures 19 and 20 and even more clearly in Figure 21, there is no meaningful difference in average quality between the two samples. In other words, the quality of decentralized reports has neither improved nor deteriorated.

Figure 21. Comparison in quality between decentralized final evaluation reports from 2013 or before, and after 2013 (n = 147)



3 | MINIMUM REQUIREMENTS FOR SYNTHETIC ANALYSIS

As a preliminary step to the next stage of the study (which consists of assessing and synthesizing information on relevance and effectiveness), we identify the evaluation reports that meet the three minimum requirements for synthetic analysis. Indeed, as laid out in Chapter 1, in order to assess the relevance and outcomes of a given programmatic area, a number of assumptions need to be met:

- The programmatic area has sufficient evaluation coverage.
- The evidence base is of sufficient quality.
- The evidence base stems (to a considerable extent) from reports commissioned by an independent evaluation office.

With regard to the first criterion, in the coverage analysis presented in Chapter 2 we mapped evaluation reports onto the respective ERs to get a sense of coverage. The results of this analysis are displayed in Chapter 2 (Figure 8).

With regard to the second criterion, in Figure 22 for each ER we present the number of final evaluation reports that received a score of 4 or more on the composite quality index (see succinct meta-evaluation presented above).

Finally, regarding the third criterion, in Figure 23 we display the number of final evaluation reports that received a score of 4 or more on the composite quality index and were commissioned by an independent evaluation office (whether UNESCO or not).

Combining these three criteria and applying the decision rule (of minimum three final evaluation reports with a quality rating of at least 4 and commissioned by an independent evaluation office) presented in the methodology section we can conclude the following. Out of 47 ERs in the 37C/5 UNESCO program, 12 (26%) meet the minimum requirements for synthetic analysis of relevance and effectiveness and 35 (74%) do not. Figure 24 summarizes this information. The detailed analysis is presented in Annex 3.

Figure 22. Distribution of final evaluation reports with a quality score of 4 or more by ER (n = 157)

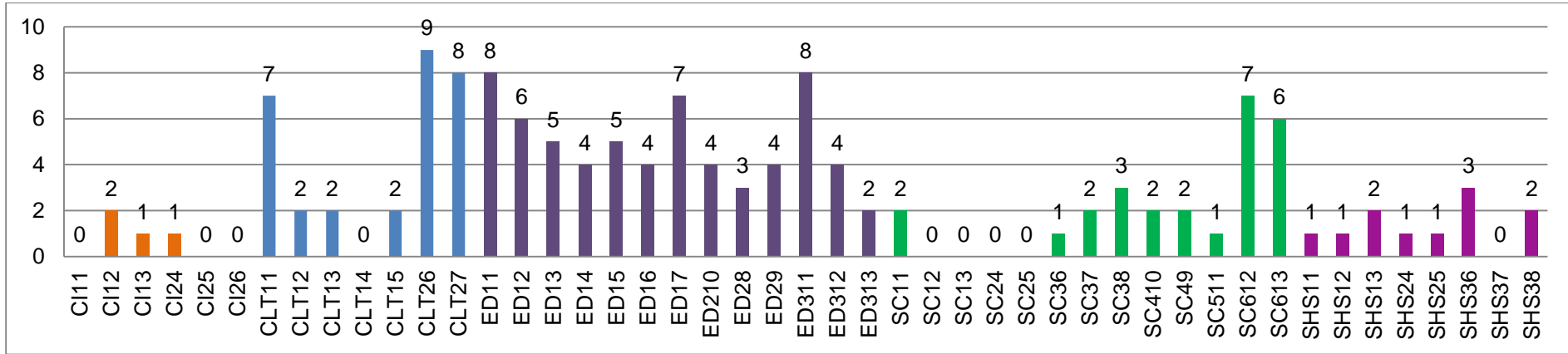


Figure 23. Distribution of final evaluation reports with a quality score of 4 or more and commissioned by an independent evaluation office by ER (n = 100)

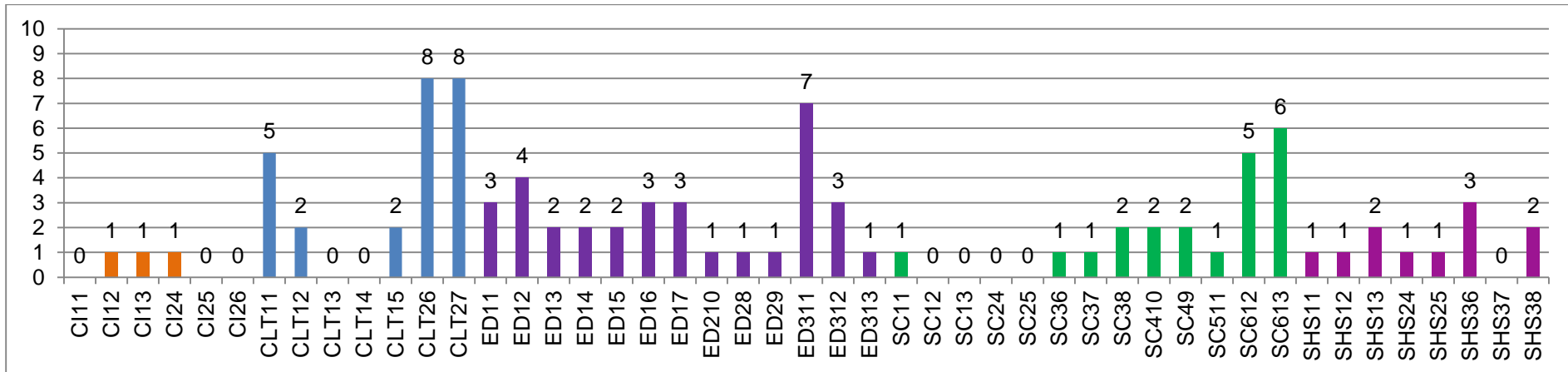
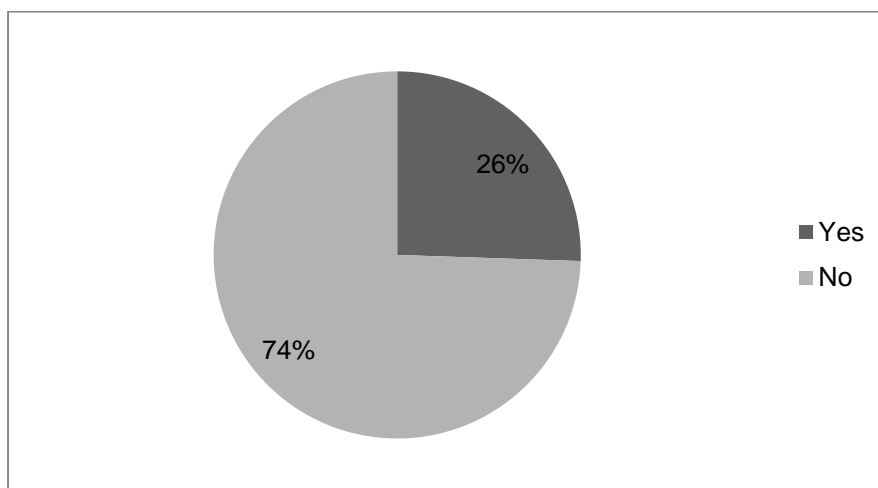


Figure 24. Total number of ERs that meet the minimum requirements for synthetic analysis (n = 47)



4 | KEY FINDINGS

On the basis of the descriptive statistical analyses of evaluation reports discussed above we can conclude the following:

Finding 3: Overall, the vast majority of evaluation reports meet the basic reporting requirements in terms of including information on implementation, output delivery, effects, and lessons learned. The evaluation reports provide abundant and rich descriptions of activities and outputs, as well as an increasingly consistent effort to draw lessons from the intervention. However, the evidence on effects (outcome or impact) is often very patchy.

Finding 4: The quality of reports has been stable over time. Notably, there has been no substantial change in the quality of reports after the 2013 Diagnostic Study of Evaluations of UNESCO's Extrabudgetary Activities.

Finding 5: The minimum requirements for assessing the relevance and effectiveness of a programmatic area of work on the basis of existing evaluations— based on the criteria of sufficient coverage, minimum quality, and independence—are met for only a fifth of UNESCO's 47 ERs. Consequently, comparative assessment of programmatic areas of work on the basis of existing evaluations is currently not possible.

CHAPTER 4: SYNTHETIC ANALYSIS OF EFFECTIVENESS AND RELEVANCE OF UNESCO WORK - ILLUSTRATIVE EVIDENCE FOR SELECTED EXPECTED RESULTS

In this chapter, we seek to illustrate what a comprehensive synthetic review would entail if the minimum requirements for synthetic analysis (coverage, quality, and independence) were met for the majority (or ideally all) of UNESCO's programmatic areas. For illustrative purposes, in Section 1 we selected five ERs (one per UNESCO sector) for which these minimum requirements were met and systematically assessed the evaluative evidence, applying the rating procedures described in Chapter 1 to assess relevance and effectiveness. For each Expected Result, we present the rating results and illustrative findings.

In Section 2, we propose a possible approach for comparative programme assessment based on multi-criteria decision analysis to complement UNESCO's quadrennial Strategic Results Report and to inform decision-making processes on strategic priorities and the strategic allocation of resources.

1 | ILLUSTRATIVE SYNTHETIC ANALYSIS OF RELEVANCE AND EFFECTIVENESS

Based on the analysis carried out in Chapter 4, we identified one ER per sector that meets the "minimum requirements for synthetic analysis", discussed in the previous chapter. We systematically applied a matrix-based assessment of relevance and effectiveness at the level of the ER. The list of criteria to assess relevance and effectiveness —although tentative— is in line with established definitions of relevance and effectiveness (e.g., OECD-DAC criteria) and can be analyzed on the basis of evaluative evidence from evaluation reports.

Generally speaking, the purpose of the analysis is not to compare with a high degree of precision the relevance and effectiveness across programmatic areas of work. However, for those ERs that meet the minimum requirements for synthetic review, the approach allows us to identify programmatic areas that are clearly underperforming and distinguish them from areas of work that are performing well on relevance and effectiveness.

The analysis was undertaken for illustrative purposes as the current evaluative evidence base did not permit a comprehensive and representative synthetic analysis of relevance and effectiveness of all of UNESCO's programmatic areas of work (in this case ERs). We reviewed the sample of evaluation reports that were mapped onto each of the 5 illustrative ERs, and applied the assessment matrices on relevance and effectiveness (described in Chapter 2). In illustrative boxes, we provide a detailed explanation of the rating procedures for some of the reports. We reviewed the following illustrative thematic areas (ERs)¹⁵:

- ED sector: Education Policy and Planning
- SC sector: Capacity for water management
- SHS sector: Youth
- CLT sector: Diversity of cultural expressions
- CI sector: Community and pluralistic media

¹⁵ The full list of ERs and their associated thematic areas can be found in Annex 1.

1. Education Policy and Planning

We found 9 evaluation reports covering the thematic area referred to as "Education Policy and Planning" of the Major Programme I (Education)¹⁶. Each report was scored based on the relevance and effectiveness criteria presented in Chapter 1. The results are presented in Table 7 and Figure 25. In Boxes 1 and 2 we provide a detailed illustration of the rating procedures for two of the reports, one illustrating an intervention with high relevance and effectiveness, and one illustrating an intervention with high relevance but more limited effectiveness.

Table 7. Assessment Matrices for Education Policy and Planning (n = 9)

RELEVANCE		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Contribution to key global agreements (e.g., MDGs, EFA)	<ul style="list-style-type: none"> There is clear evidence that an intervention is well-aligned with a global agreement Dichotomous variable: yes versus no 	3/9
Response to UNESCO's core priorities	<ul style="list-style-type: none"> There is clear evidence that an intervention has remained relevant to UNESCO's core priorities (as expressed in current C/4 document) Dichotomous variable: yes versus no 	4/9
Alignment with national/regional initiatives	<ul style="list-style-type: none"> There is clear evidence that an intervention has developed synergies with related initiatives at the country/regional level Dichotomous variable: yes versus no 	7/9
Response to clearly defined needs of the targeted beneficiaries	<ul style="list-style-type: none"> There is clear evidence that an intervention responds to a well-targeted national or local need Dichotomous variable: yes versus no 	5/9
Uniqueness/Niche	<ul style="list-style-type: none"> There is clear evidence that an intervention embodies UNESCO's specificities and highlight its niche in a particular domain Dichotomous variable: yes versus no 	4/9
RELEVANCE	<ul style="list-style-type: none"> Composite variable, proxy of relevance of an intervention summarizing information on variables <ul style="list-style-type: none"> - contribution to global priorities - response to UNESCO's core mandate - addressing a specific (national/local) need - synergy with related initiatives at the country/regional level - uniqueness/niche Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on the 5 dimensions 5/9

¹⁶ See Annex 1 for a complete list of ERs.

EFFECTIVENESS		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Output delivery	<ul style="list-style-type: none"> There is clear evidence that the intervention delivered most planned outputs Dichotomous variable: yes versus no 	9/9
Outcome achievement	<ul style="list-style-type: none"> There is clear evidence that the intervention is likely to achieve intended (or unintended) positive outcomes Dichotomous variable: yes versus no 	4/9
Catalytic role and effects	<ul style="list-style-type: none"> There is clear evidence that an intervention is likely to have a catalytic effect Dichotomous variable: yes versus no 	5/9
Sustainability of partnerships	<ul style="list-style-type: none"> There is clear evidence that the intervention is supported by a cooperation with partners Dichotomous variable: yes versus no 	1/9
Sustainability of effects	<ul style="list-style-type: none"> There is clear evidence that the intervention will lead to long-term sustained change Dichotomous variable: yes versus no 	3/9
EFFECTIVENESS	<ul style="list-style-type: none"> Composite variable, proxy of the effectiveness of an intervention summarizing information on variables: <ul style="list-style-type: none"> - output delivery - outcome achievement - catalytic effect - sustainability - scale Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 3/9

Box 1. Example of an Education Policy and Planning project with high relevance and effectiveness

Educate a Child in Iraq: Phase I (Evaluation completed in April 2015)

The project's primary objective was to bring 30,000 out of school children from rural areas (with gender parity among students) into schools through the application of the Accelerated Learning Programme (ALP). The project was deemed relevant on four of the five relevance criteria: enhancing enrolment and retention of out of school children is at the core of the Education for All global agreement and MDG2 of "achieving universal primary education" as well as of UNESCO's strategic objectives in the Education sector. The application of the ALP methods shows alignment with UNESCO's focal areas. In addition, the project responded to one of the government of Iraq's main priorities of reconstructing and rehabilitating schools in conflict areas. Finally, the project was intended to contribute to specific local needs in well-defined areas of the country, namely Baghdad, Basra, Ninewa, and Erbil. However, the evaluation report does not present evidence that UNESCO occupied a particular niche. This is not to say that UNESCO did not have a particular comparative advantage, but that there was no discussion of other intervening actors or of UNESCO's comparative strength in the evaluation report.

With regards to effectiveness, the project scored well on three of the five criteria. The project exceeded its primary targets by enrolling more than 37,000 students (initial target 30,000) and constructing or renovating 110 schools (initial target: 100). Beneficiary surveys and interviews with officials in Iraq's Ministry of Education showed a positive perception of the project from the main stakeholders. It also

made progress towards the second phase of the project by designing a scaling-up strategy and setting up a well-needed Education Management Information System (EMIS). With regards to progress towards outcomes, the evaluation showed mixed results. On the one hand, 70% of the enrolled students moved on to the next level and were retained in school, which is considered a largely positive outcome. On the other hand, there were clear shortcomings in terms of the provision of the right number of teachers, learning spaces, and meeting the basic needs of quality education, such as the provision of stationery, curriculum, and textbooks. Importantly, the assessment of learning performance was not part of the project design, as the project did not address issues of teaching-learning effectiveness, limiting the measure of success to the number of enrolled learners in the schools. The project was successful in engaging multiple partners, including the Iraqi Ministry of Education. It also fostered synergies with other UNESCO-led projects supported by a range of contributors, such as UNHCR, the Kingdom of Saudi Arabia, and the Emergency Response Fund. While these strong partnerships ensure a certain degree of continuity, the report does not provide evidence of sustainability of effects beyond the fact that a phase II of the operation was scheduled. Finally, there was no evidence of catalytic effects, which is partly related to the nature of the project, with little focus on capacity building and systems institutionalization (the EMIS being one of the notable exceptions).

Box 2. Example of an Education Policy and Planning intervention with high relevance and limited (evidence of) effectiveness

Review of the UNESCO Institute for Lifelong Learning (UIL) (Evaluation completed in March 2013)

The institute scored positively on four of the five relevance criteria. Under the umbrella of lifelong learning policy and practice, and with its particular focus on literacy, the institute is undeniably of continued relevance to UNESCO's Education sector strategy and to global objectives within the EFA and the SDG frameworks. Moreover, the report delineates niche areas of the institute, unmatched by other actors in the sector. For instance, the report notes: *"UIL is recognized as the only global Institute bringing a comparative perspective in the specialized areas of its mandate and for raising the profile of adult education on national education agendas."* In addition, the capacity to link stakeholders at the national level while providing a global perspective is considered a clear comparative advantage of the institute in the area of Adult Education. However, given its strategic position, working primarily at the global and/or regional level, the institute is not set up to address specific local needs.

With regards to effectiveness, the institute scores well on two out of the five indicators. UIL has produced many important outputs in all of its main functions (e.g., capacity building, research). In addition, the evaluation highlights the particular strength of the institute in terms of catalyzing international actions through a potent networking capacity. However, the report notes that little is known about the outcomes of the institute's activities, including on probably catalytic effects. In addition, the evaluation highlights a number of missed opportunities for partnerships, in particular within the UNESCO "Education Family". UIL is poorly integrated into the Education sector strategy and there is limited cooperation with other Category I institutes. The sustainability of the institute was analyzed solely through the prism of financial sustainability and some concerns were raised with regards to the staffing model given the falling regular programme budget allocated to the institute. It should be noted that the relatively low score on effectiveness in this case is much the result of the limited focus of the evaluation rather than the reality of the intervention.

2. Capacity for water management

There were 8 reports covering the thematic area referred to as "Capacity for Water Management" of the Major Programme II (Natural Sciences). Each report was scored based on the relevance and effectiveness criteria presented in Chapter 1. The results are presented in Table 8 and Figure 26. In

Boxes 3 and 4 we provide a detailed illustration of the rating procedures for two of the reports, one illustrating an intervention with high relevance and effectiveness, and one illustrating an intervention with high relevance but more limited effectiveness.

Table 8. Assessment Matrices for Capacity for water management (n = 8 reports)

RELEVANCE		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Contribution to key global agreements (e.g., MDGs, EFA)	<ul style="list-style-type: none"> There is clear evidence that an intervention is well-aligned with a global agreement Dichotomous variable: yes versus no 	6/8
Response to UNESCO's core priorities	<ul style="list-style-type: none"> There is clear evidence that an intervention has remained relevant to UNESCO's core priorities (as expressed in current C/4 document) Dichotomous variable: yes versus no 	5/8
Alignment with national/regional initiatives	<ul style="list-style-type: none"> There is clear evidence that an intervention has developed synergies with related initiatives at the country/regional level Dichotomous variable: yes versus no 	4/8
Response to clearly defined needs of the targeted beneficiaries	<ul style="list-style-type: none"> There is clear evidence that an intervention responds to a well-targeted national or local need Dichotomous variable: yes versus no 	4/8
Uniqueness/Niche	<ul style="list-style-type: none"> There is clear evidence that an intervention embodies UNESCO's specificities and highlight its niche in a particular domain Dichotomous variable: yes versus no 	3/8
RELEVANCE	<ul style="list-style-type: none"> Composite variable, proxy of relevance of an intervention summarizing information on variables <ul style="list-style-type: none"> - contribution to global priorities - response to UNESCO's core mandate - addressing a specific (national/local) need - synergy with related initiatives at the country/regional level - uniqueness/niche Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 4/8
EFFECTIVENESS		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Output delivery	<ul style="list-style-type: none"> There is clear evidence that the intervention delivered most planned outputs Dichotomous variable: yes versus no 	7/9
Outcome achievement	<ul style="list-style-type: none"> There is clear evidence that the intervention is likely to achieve intended (or unintended) positive outcomes Dichotomous variable: yes versus no 	3/9
Catalytic role and effects	<ul style="list-style-type: none"> There is clear evidence that an intervention is likely to have a catalytic effect Dichotomous variable: yes versus no 	7/9

Sustainability of partnerships	<ul style="list-style-type: none"> • There is clear evidence that the intervention is supported by a cooperation with partners • Dichotomous variable: yes versus no 	3/9
Sustainability of effects	<ul style="list-style-type: none"> • There is clear evidence that the intervention will lead to long-term sustained change • Dichotomous variable: yes versus no 	5/9
EFFECTIVENESS	<ul style="list-style-type: none"> • Composite variable, proxy of the effectiveness of an intervention summarizing information on variables: <ul style="list-style-type: none"> - output delivery - outcome achievement - catalytic effect - sustainability - scale • Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 6/8

Box 3. Example of a joint Water Management intervention with high relevance and effectiveness

Evaluation of Adaptation to Climate Change to Sustain Jordan's MDG Achievement (Evaluation completed in March 2013)

The Joint Programme (JP) was intended to support adaptations in the water and public health sectors, in particular through an upgrade of the national drinking water safety management system and to enhance drinking water security in response to water scarcity.

The intervention scored well on all five "relevance" criteria. It was deemed particularly relevant in relation to the global agenda for addressing climate change and more particularly achieving MDG-7. The report highlights the strong alignment of the JP with UNESCO's IHP strategy and the mandate of UNESCO's wider 'Water Family'. In addition, the report concluded that the JP was highly relevant in developing a climate change adaptation agenda at the national level in Jordan, which was inexistent prior to the intervention. At the local level, the project was well-targeted to address the needs of specific communities suffering from water scarcity. Finally, the evaluators explicitly lay out IHP's unique contribution to the intervention, articulating the Programme's niche areas in the following terms: *"IHP is the only broadly-based science programme of the UN system in the area of water management."* The report also highlighted the embeddedness of IHP's activities in Jordan, with an IHP committee that was established in the country in 1992.

With regard to effectiveness, the JP scored well on all five dimensions. First, the JP delivered on all of the planned outputs. Tangible outcomes, including positive unintended outcomes, were evidenced in the report. The evaluators highlighted a number of institutional externalities of the JP that would ensure the pursuit of some of the activities by the Jordanian government. The report highlights the development of 6 climate change adaptation strategies: in the Ministry of Health, incorporated in the new National Health Strategy finalized in 2013; in the development process of the soon-to-be finalized National Climate Change Policy; the creation of a Directorate on Climate Change at the Ministry of Environment; and the creation by a Ministerial Order of a committee on water quality surveillance chaired by the Ministry of Health. Moreover, the JP achieved some important catalytic effects. For example, it supported the implementation of Water Safety Plans as a preventative risk management framework by five utilities serving 85% of the Jordanian population. The JP's emphasis on capacity-building and institutionalization, and the high ownership of the intervention by the Government of Jordan, led the evaluators to conclude that the effects of the intervention were likely to be sustained over time.

Box 4. Example of a Water Management intervention with high relevance and low effectiveness

Evaluation of the Regional Center for Training and Water Studies in Arid and Semi-Arid Zones (RCTWS-Egypt): a UNESCO Category II Institute (Evaluation completed in June 2014)

The center scores well on three of the five relevance indicators. The center's mandate and strategy of fostering training and research on water in arid and semi-arid zones is well-aligned with the post-2015 agenda on sustainable development, and fits well in UNESCO IHP's strategy. Moreover, the evaluation established that it is an important structure to address the training needs of the region, based on consultation and needs assessment in the region. However, the center does not occupy a niche as there are already a number of intervening actors, including within the University of Cairo, and a National center for research on water, which occupy the water studies space. Finally, the center does not directly focus on addressing local needs.

In terms of effectiveness, the center scores well on one of the indicators. Due to shortage of funds, a drop in demand, political and economic instability in the country and a vacant regional hydrologist position in the UNESCO Cairo Office, many of the planned activities did not take place. Only a small share of the planned training took place with a limited number of trainees. While between 2004 and 2008, the center trained on average 300 regional participants, between 2009 and 2014, this average dropped considerably to about 80 trainees. On the other hand, the center has generated a number of publications related to IHP VII. The evaluation was unable to measure progress at the level of outcomes. While the center originally managed to foster quality collaborations with the government of Egypt and a number of international institutions, including with the Japan International Cooperation Agency and the European Union, the evaluation notes that the center is continuing to face important challenges, in terms of funding, governance structure, lack of a clear institutional strategy and technical steering committee. These challenges threaten the continuity and sustainability of the institute's activities.

Figure 25. Education Policy and Planning: Relevance and Effectiveness (n = 9)

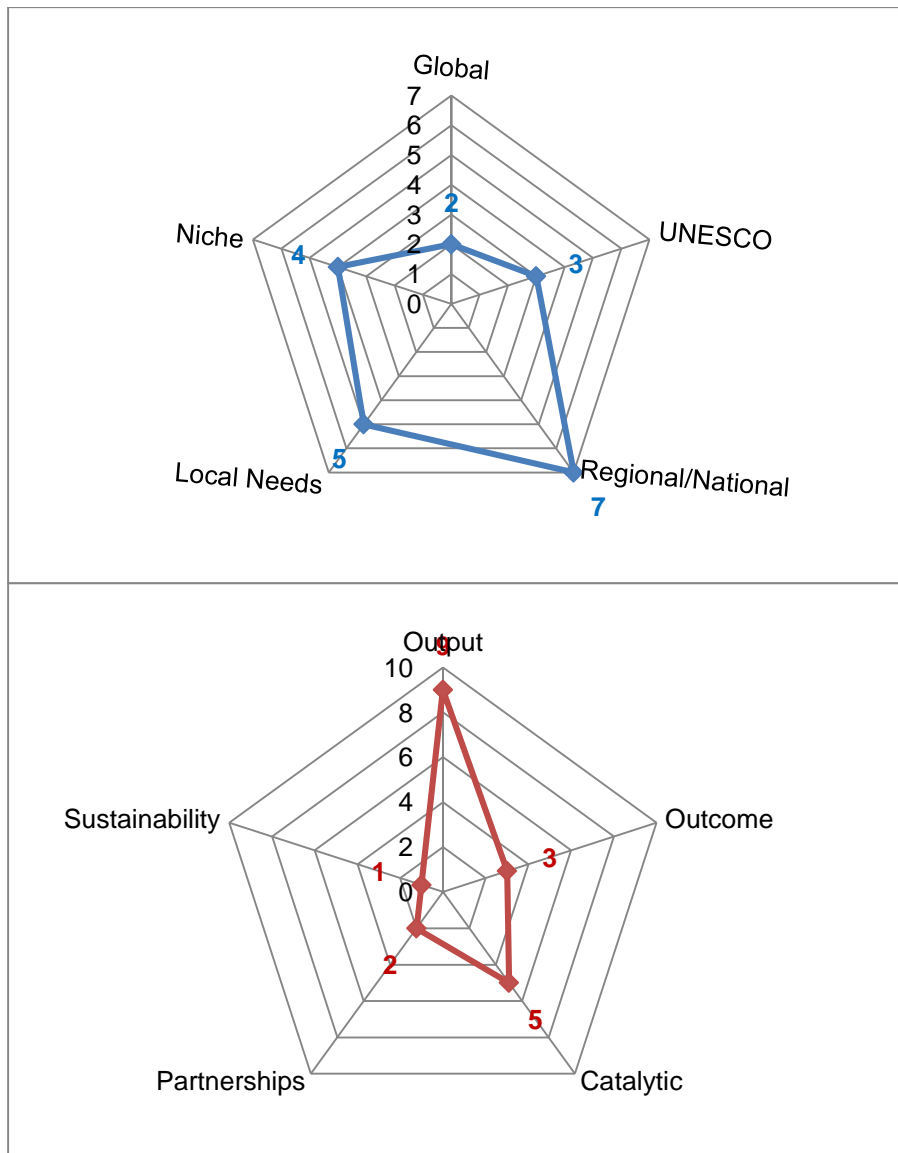
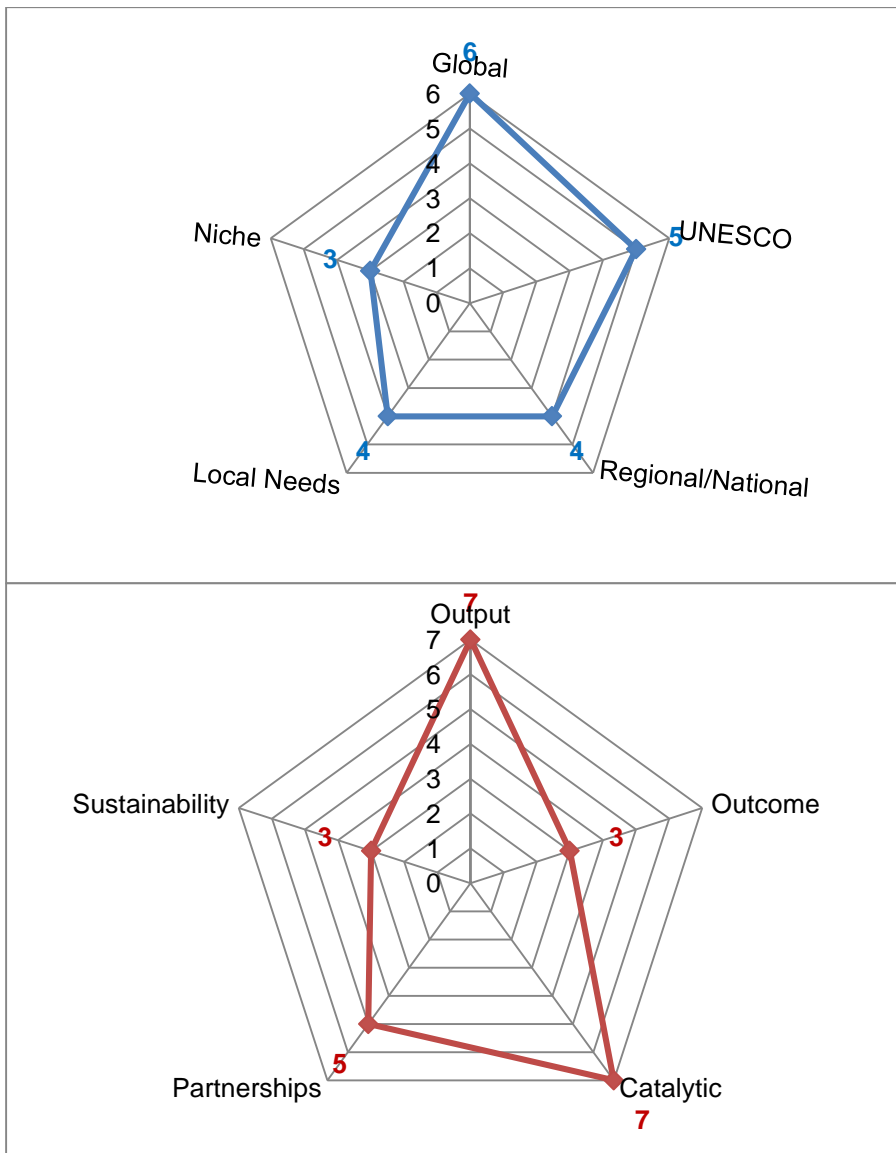


Figure 26. Capacity for water management: Relevance and Effectiveness (n = 9)



3. Youth¹⁷

There were 5 reports covering the thematic area referred to as "Youth" of the Major Programme III (Social and Human Sciences). Each report was scored based on the relevance and effectiveness criteria presented in Chapter 1. The results are presented in Table 9 and displayed in Figure 27. In Box 5 we provide a detailed illustration of the rating procedures for an intervention with high relevance and effectiveness.

Table 9. Assessment Matrices for Youth (n = 5 reports)

RELEVANCE		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Contribution to key global agreements (e.g., MDGs, EFA)	<ul style="list-style-type: none"> There is clear evidence that an intervention is well-aligned with a global agreement Dichotomous variable: yes versus no 	3/5
Response to UNESCO's core priorities	<ul style="list-style-type: none"> There is clear evidence that an intervention has remained relevant to UNESCO's core priorities (as expressed in current C/4 document) Dichotomous variable: yes versus no 	0/5
Alignment with national/regional initiatives	<ul style="list-style-type: none"> There is clear evidence that an intervention has developed synergies with related initiatives at the country/regional level Dichotomous variable: yes versus no 	5/5
Response to clearly defined needs of the targeted beneficiaries	<ul style="list-style-type: none"> There is clear evidence that an intervention responds to a well-targeted national or local need Dichotomous variable: yes versus no 	5/5
Uniqueness/Niche	<ul style="list-style-type: none"> There is clear evidence that an intervention embodies UNESCO's specificities and highlight its niche in a particular domain Dichotomous variable: yes versus no 	0/5
RELEVANCE	<ul style="list-style-type: none"> Composite variable, proxy of relevance of an intervention summarizing information on variables <ul style="list-style-type: none"> -contribution to global priorities - response to UNESCO's core mandate - addressing a specific (national/local) need - synergy with related initiatives at the country/regional level - uniqueness/niche Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 3/5
EFFECTIVENESS		

¹⁷ A number of additional caveats for this assessment need to be mentioned. There were 8 evaluations covering the programmatic area of Youth, 5 of which were final evaluations and 3 mid-term evaluations. Moreover, most of the final evaluations were from the MDG-F and covered UNESCO's role only superficially. All of the evaluations took place prior to 2013. Finally, while the theme of youth is the focus of the evaluation, most of the work described is "intersectoral" in nature. For instance, many interventions involve strengthening TVET for migrant youth.

Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Output delivery	<ul style="list-style-type: none"> • There is clear evidence that the intervention delivered most planned outputs • Dichotomous variable: yes versus no 	5/5
Outcome achievement	<ul style="list-style-type: none"> • There is clear evidence that the intervention is likely to achieve intended (or unintended) positive outcomes • Dichotomous variable: yes versus no 	3/5
Catalytic role and effects	<ul style="list-style-type: none"> • There is clear evidence that an intervention is likely to have a catalytic effect • Dichotomous variable: yes versus no 	3/5
Sustainability of partnerships	<ul style="list-style-type: none"> • There is clear evidence that the intervention is supported by a cooperation with partners • Dichotomous variable: yes versus no 	2/5
Sustainability of effects	<ul style="list-style-type: none"> • There is clear evidence that the intervention will lead to long-term sustained change • Dichotomous variable: yes versus no 	3/5
EFFECTIVENESS	<ul style="list-style-type: none"> • Composite variable, proxy of the effectiveness of an intervention summarizing information on variables: <ul style="list-style-type: none"> - output delivery - outcome achievement - catalytic effect - sustainability - scale • Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 3/5

Box 5. Example of a Youth joint intervention with moderately high relevance and high effectiveness

Evaluation MDG-F Programa Conjunto de Juventud, Empleo y Migración: Una Ventanilla Única para el empleo juvenil en Desamparados y Upala: Costa Rica (Evaluation completed in January 2013)

The Joint Programme (JP) aimed at creating opportunities for young migrants either by ensuring their enrollment in school or a training programme or by enhancing their job preparedness.

The intervention scored well on three of the five relevance criteria. The JP addressed a direct priority of Costa Rica's government and an overarching global priority. In the design phase, the JP made serious efforts to bring the programme to very remote areas with the highest rate of unemployment and weak socio-economic indicators, thereby matching the needs of the communities most in want of such type of interventions. No discussion of the relevance of the intervention for UNESCO was included in the report.

With regard to effectiveness, the intervention also scored well on four of the five dimensions. First, the JP delivered on all of its planned outputs, including in terms of identifying, registering and orienting young migrants with a personal development plan. A number of positive outcomes of the programme were also measured in the evaluation, including more than half of the 3,400 young migrants reached through the JP receiving technical training, and 10% already having obtained formal employment.

Moreover, the JP has had a number of catalytic effects. For example, the evaluators highlighted the publication of an ABC of youth migrant employment, which has gained traction in other cantons. The JP intently worked towards strengthening existing partnerships and networks on youth employment and resulted in the elaboration of an integrated local strategy with clear division of roles and responsibilities among the intervening partners, all the while ensuring local ownership. Consequently, the JP's effects were deemed likely to be sustained after the closure of the programme.

4. Diversity of cultural expressions

There were 8 reports covering the thematic area referred to as "Diversity of cultural expressions" of the Major Programme IV (Culture). Each report was scored based on the relevance and effectiveness criteria presented in Chapter 1. The results are presented in Table 10 and displayed in Figure 28. In Boxes 6 and 7 we provide a detailed illustration of the rating procedures for two interventions, one with high relevance and effectiveness, and one with high relevance but more limited effectiveness.

Table 10. Assessment Matrices for Diversity of cultural expressions (n = 8 reports)

RELEVANCE		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Contribution to key global agreements (e.g., MDGs, EFA)	<ul style="list-style-type: none"> There is clear evidence that an intervention is well-aligned with a global agreement Dichotomous variable: yes versus no 	7/8
Response to UNESCO's core priorities	<ul style="list-style-type: none"> There is clear evidence that an intervention has remained relevant to UNESCO's core priorities (as expressed in current C/4 document) Dichotomous variable: yes versus no 	8/8
Alignment with national/regional initiatives	<ul style="list-style-type: none"> There is clear evidence that an intervention has developed synergies with related initiatives at the country/regional level Dichotomous variable: yes versus no 	7/8
Response to clearly defined needs of the targeted beneficiaries	<ul style="list-style-type: none"> There is clear evidence that an intervention responds to a well-targeted national or local need Dichotomous variable: yes versus no 	2/8
Uniqueness/Niche	<ul style="list-style-type: none"> There is clear evidence that an intervention embodies UNESCO's specificities and highlight its niche in a particular domain Dichotomous variable: yes versus no 	3/8
RELEVANCE	<ul style="list-style-type: none"> Composite variable, proxy of relevance of an intervention summarizing information on variables <ul style="list-style-type: none"> -contribution to global priorities - response to UNESCO's core mandate - addressing a specific (national/local) need - synergy with related initiatives at the country/regional level <ul style="list-style-type: none"> - uniqueness/niche Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 7/8
EFFECTIVENESS		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"

Output delivery	<ul style="list-style-type: none"> • There is clear evidence that the intervention delivered most planned outputs • Dichotomous variable: yes versus no 	6/8
Outcome achievement	<ul style="list-style-type: none"> • There is clear evidence that the intervention is likely to achieve intended (or unintended) positive outcomes • Dichotomous variable: yes versus no 	2/8
Catalytic role and effects	<ul style="list-style-type: none"> • There is clear evidence that an intervention is likely to have a catalytic effect • Dichotomous variable: yes versus no 	5/8
Sustainability of partnerships	<ul style="list-style-type: none"> • There is clear evidence that the intervention is supported by a cooperation with partners • Dichotomous variable: yes versus no 	3/8
Sustainability of effects	<ul style="list-style-type: none"> • There is clear evidence that the intervention will lead to long-term sustained change • Dichotomous variable: yes versus no 	0/8
EFFECTIVENESS	<ul style="list-style-type: none"> • Composite variable, proxy of the effectiveness of an intervention summarizing information on variables: <ul style="list-style-type: none"> - output delivery - outcome achievement - catalytic effect - sustainability - scale • Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 4/8

Box 6. Example of a Diversity of Cultural Expressions intervention with high relevance and effectiveness

Evaluation of the 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions (Evaluation completed in April 2014)

The convention scored well on four of the five relevance dimensions. The convention is deemed particularly relevant to the post-2015 agenda with cultural industries recognized as a key lever of sustainable development; it is also core to UNESCO's CLT sector strategy. Given the increase in ratification of the convention among countries, the convention has undeniable national traction. However, little can be extrapolated from the report with regards to relevance at the local level. Finally, the report clearly delineates the unique role that UNESCO plays in ensuring the translation of the convention into policy and practice, with a clear niche contribution within the UN.

The convention also scored well on four of the five effectiveness indicators. UNESCO's role is seen as particularly positive in establishing an enabling environment for the convention, notably through the provision of a space for exchange of experiences, the provision of technical assistance, and the consolidation of a knowledge base for cultural policies. Moreover, the evaluation concludes that the 2005 Convention has had a positive impact in some policy areas and in some regions. Notable examples include, strengthening cultural policies and legislation in several countries in Africa, Latin America and Asia, through the setting-up of specialized ministries or other governmental departments, as well as more specific policies and programmes in the field of cultural and creative industries. Nevertheless, the evaluation also underscores the uneven outcome pictures resulting from the convention: "*Some excellent examples exist, while other countries have so far failed to take any significant steps forward.*" Interesting instances of catalytic effects of the convention are mentioned in the report. For example, in a number of countries, the ratification and coming into force of the convention has given the necessary

impetus to set up new ministries, policy departments or agencies concerned with cultural policy and cultural industries. Similarly, the Convention has been mentioned as a source of inspiration for new legislation or national policies in the field of culture. Moreover, in a number of countries the convention has played a critical role as a convening platform for several agencies and partners to come together and support newly adopted cultural policy.

Box 7. Example of a joint Diversity of Cultural Expressions intervention with moderately high relevance and limited effectiveness

Evaluation of the Joint Programme "Patrimoine, Tradition Et Créativité Au Service Du Développement Durable " in Mauritania (Evaluation completed in July 2013)

The joint programme (JP) scored well on three of the five relevance indicators. Apart from being well aligned with global priorities on sustainable development and the core priorities of each partner agencies, it was also deemed highly relevant at the national level in the valorization of Mauritania's cultural heritage. It also responded to the government's priority of reducing poverty and improving the income of professionals of the CLT sector, notably crafts and tourism, with a particular emphasis on youth and women.

On the other hand, the JP scores low on effectiveness: the programme did not succeed in delivering on its planned outputs. Many planned activities did not take place for lack of funding. Moreover, the report notes that the lack of a proper M&E system hinders the assessment and tracking of any potential outcome resulting from the activities that did take place. The ownership by the national authorities and regional stakeholders is deemed quite low which casts a doubt on the potential for sustaining the efforts. No specific catalytic effects of the programme were highlighted in the report. These weaknesses were attributed to a lack of clarity in the implementation strategy, the weak coordination among the members of the JP, the absence of a proper M&E system that would allow to track progress and identify challenges along the way, and the limited funding available to carry out the ambitious plan of activities.

Figure 27. Youth: Relevance and Effectiveness (n = 5)

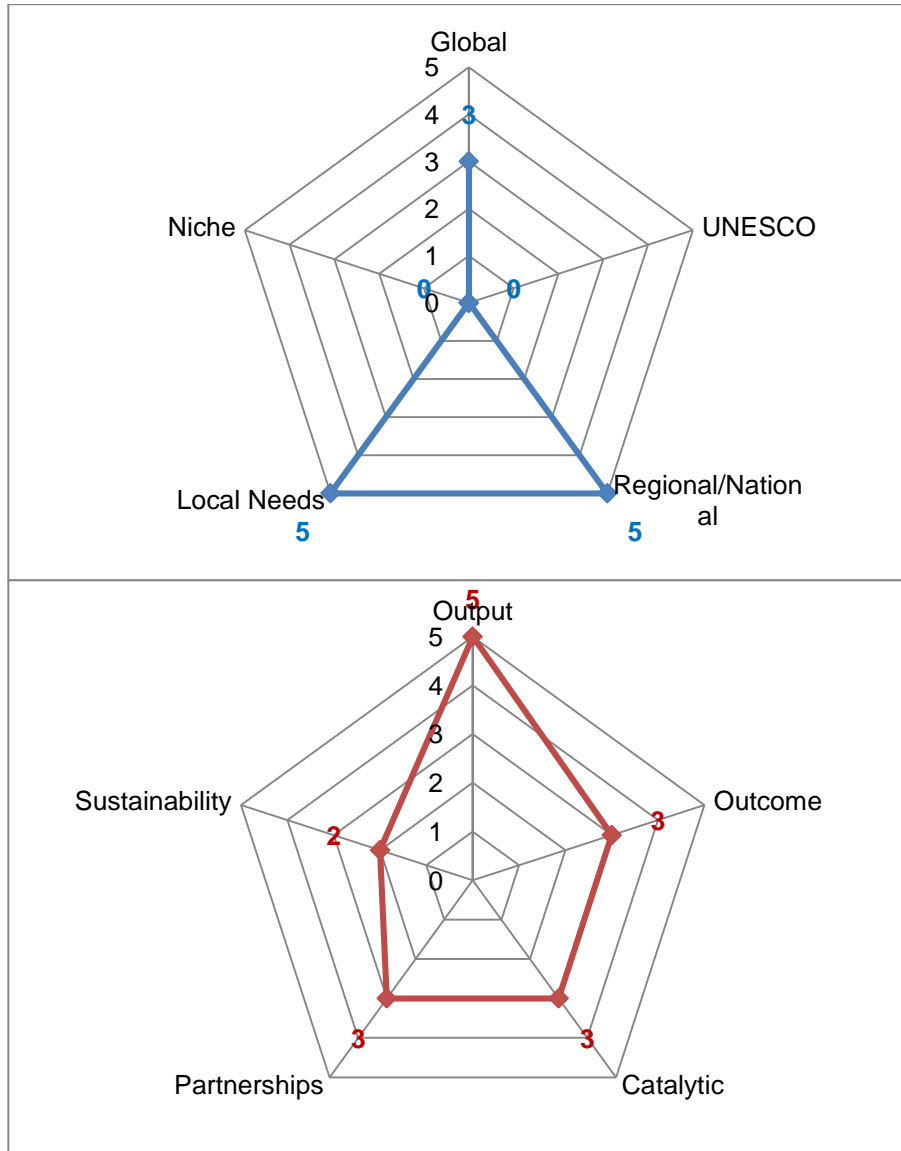
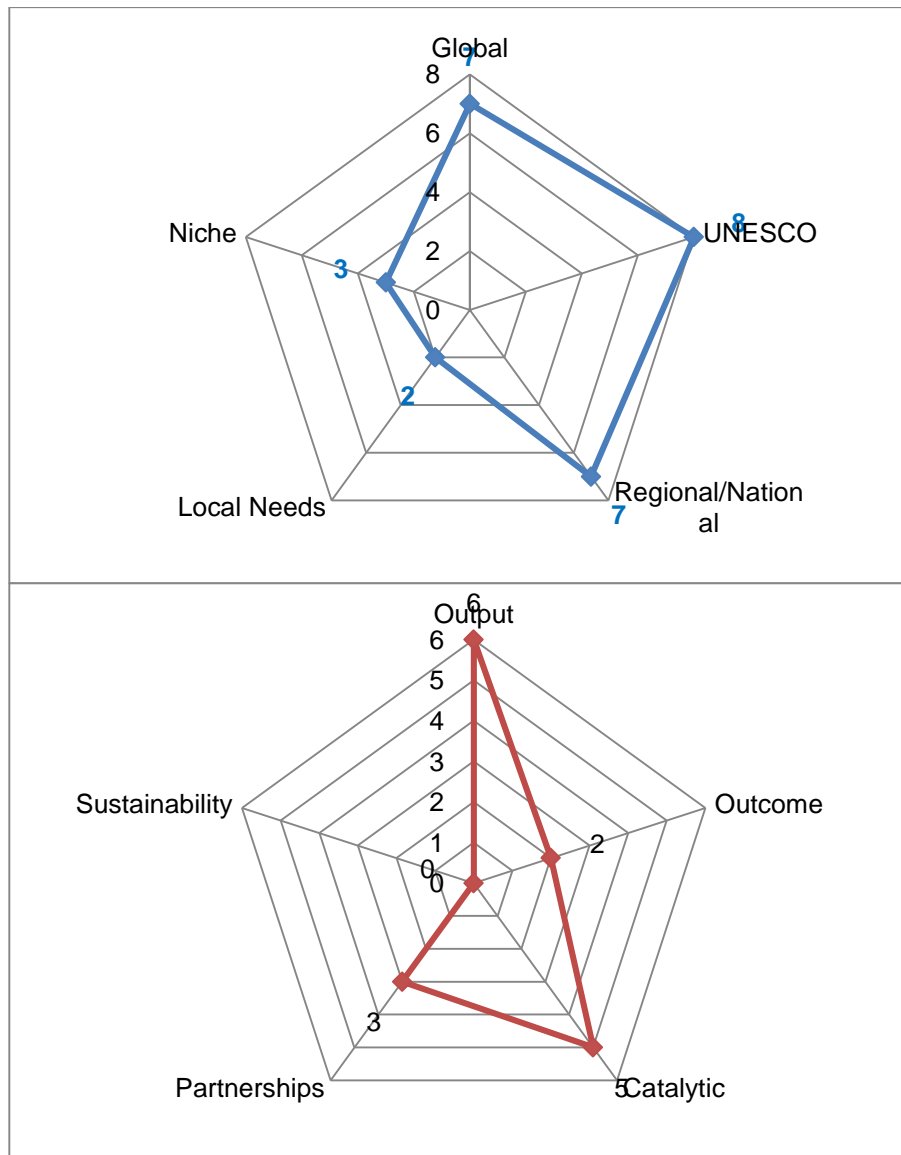


Figure 28. Diversity of cultural expressions: Relevance and Effectiveness (n = 8)



5. Community and pluralistic media

There were only 3 reports covering the thematic area referred to as "Community and pluralistic media" of the Major Programme V (Communication and Information). Each report was scored based on the relevance and effectiveness criteria presented in Chapter 1. The results are presented in Table 11 and displayed in Figure 29. In Box 8 we provide a detailed illustration of the rating procedures for an intervention with low relevance but high effectiveness.

Table 11. Assessment Matrices for Community and pluralistic media (n = 3 reports)

RELEVANCE		
Criteria	Definitions and scales	Ratio of reports showing evidence of "yes"
Contribution to key global agreements (e.g., MDGs, EFA)	<ul style="list-style-type: none"> There is clear evidence that an intervention is well-aligned with a global agreement Dichotomous variable: yes versus no 	0/3
Response to UNESCO's core priorities	<ul style="list-style-type: none"> There is clear evidence that an intervention has remained relevant to UNESCO's core priorities (as expressed in current C/4 document) Dichotomous variable: yes versus no 	1/3
Alignment with national/regional initiatives	<ul style="list-style-type: none"> There is clear evidence that an intervention has developed synergies with related initiatives at the country/regional level Dichotomous variable: yes versus no 	2/3
Response to clearly defined needs of the targeted beneficiaries	<ul style="list-style-type: none"> There is clear evidence that an intervention responds to a well-targeted national or local need Dichotomous variable: yes versus no 	2/3
Uniqueness/Niche	<ul style="list-style-type: none"> There is clear evidence that an intervention embodies UNESCO's specificities and highlight its niche in a particular domain Dichotomous variable: yes versus no 	1/3
RELEVANCE	<ul style="list-style-type: none"> Composite variable, proxy of relevance of an intervention summarizing information on variables <ul style="list-style-type: none"> - contribution to global priorities - response to UNESCO's core mandate - addressing a specific (national/local) need - synergy with related initiatives at the country/regional level - uniqueness/niche Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 0/3
EFFECTIVENESS		
Criteria	Definitions and Scales	Ratio of reports showing evidence of "yes"
Output delivery	<ul style="list-style-type: none"> There is clear evidence that the intervention delivered most planned outputs Dichotomous variable: yes versus no 	3/3
Outcome achievement	<ul style="list-style-type: none"> There is clear evidence that the intervention is likely to achieve intended (or unintended) positive outcomes Dichotomous variable: yes versus no 	1/3
Catalytic role and effects	<ul style="list-style-type: none"> There is clear evidence that an intervention is likely to have a catalytic effect Dichotomous variable: yes versus no 	1/3
Sustainability of partnerships	<ul style="list-style-type: none"> There is clear evidence that the intervention is supported by a cooperation with partners Dichotomous variable: yes versus no 	1/3

Sustainability of effects	<ul style="list-style-type: none"> • There is clear evidence that the intervention will lead to long-term sustained change • Dichotomous variable: yes versus no 	0/3
EFFECTIVENESS	<ul style="list-style-type: none"> • Composite variable, proxy of the effectiveness of an intervention summarizing information on variables: <ul style="list-style-type: none"> - output delivery - outcome achievement - catalytic effect - sustainability - scale • Ordinal variable ranging from 0 to 5 	Ratio of reports scoring 3 or more on 5 dimensions 2/3

Box 8. Example of a Community and Plurality of Media joint intervention with low relevance and high effectiveness

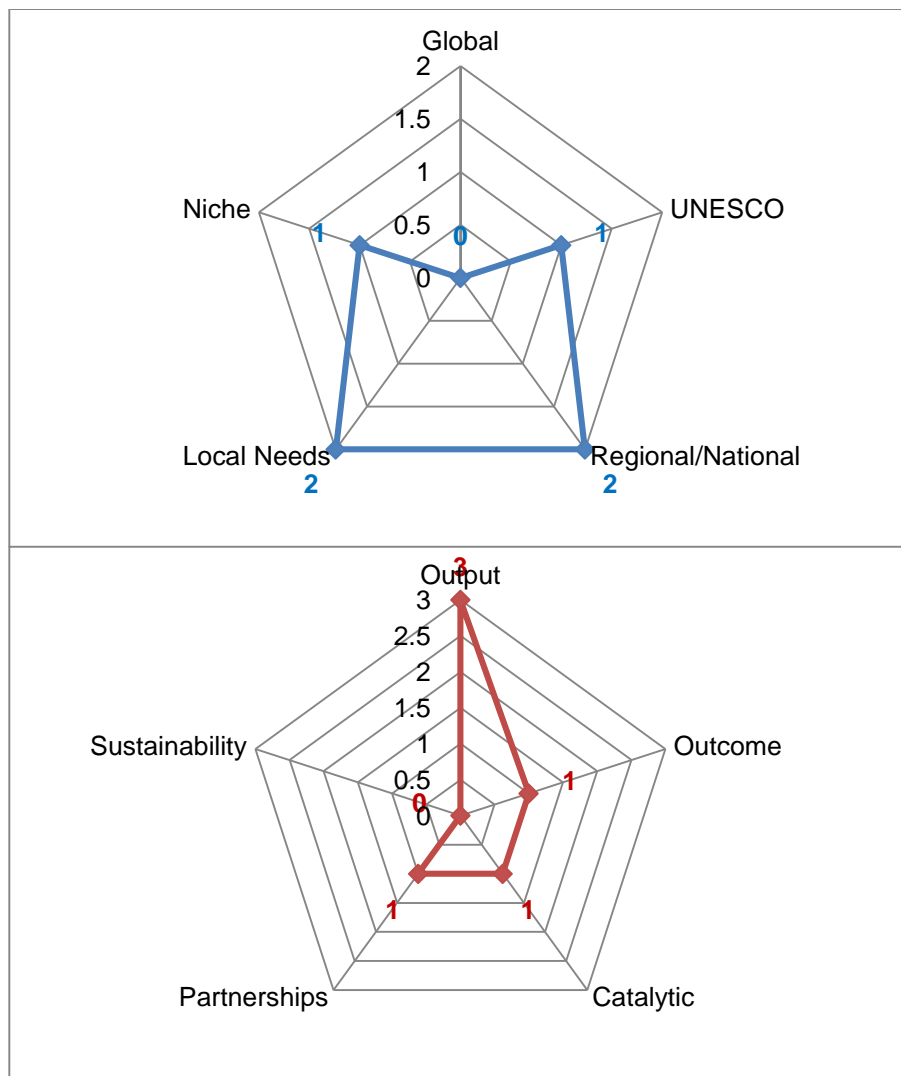
Evaluation of "Comprehensive Community Services to Improve Human Security for the Rural Disadvantaged Populations in Mongolia" (Evaluation completed in May 2012).

UNESCO was involved in this comprehensive joint intervention funded by the UN Trust Fund for Human Security. UNESCO was the project lead agency, and was also in charge of one of the five overarching goals: "improving access to information among rural populations, especially ethnic/linguistic minorities." The assessment of relevance and effectiveness focused on that particular part of the intervention.

The project scored 2 out of 5 on the relevance criteria. While the project was broadly aligned with UNESCO's communication and information sector strategy to enhance access to information, especially of minorities and rural population, and is also loosely aligned with a number of global human rights commitments, the report does not provide explicit evidence of the project's alignment with any global or national/regional agendas and priorities. Moreover, the comparative advantage of UNESCO to undertake this downstream work was not established. On the other hand, the evaluation recognizes the involvement of the Mongolian civil society organizations in the needs assessment and establishment of the community radios, which enabled this strand of the project to respond to well-defined local needs.

In terms of effectiveness, the project component on access to information scored well on three of the five effectiveness indicators. The evaluation was able to verify that all of the major planned outputs were delivered, including training for local journalists, anchors and translators; provide a framework for the registration of community radios as NGOs; and the continued subscription of the satellite provider for the second satellite TV channel of Mongolia National Broadcaster. The evaluation report attests to some noteworthy improvements in access of disadvantaged populations to information and culturally relevant programmes through the community radios and the TV channel, that were supported by the project. In addition, the project benefited from high ownership at the local level. However, the evaluation notes that long-term sustainability of project outcomes is uncertain. A number of vulnerabilities to the sustained effects of community radios are highlighted in the report, including the challenges in generating funds, the need to continuously meet the expectations of local audiences, and the technical maintenance of equipment.

Figure 29. Community and pluralistic media: Relevance and Effectiveness (n = 3)



2| THE CONTRIBUTION OF EVALUATIONS TO A COMPARATIVE PROGRAMME ASSESSMENT

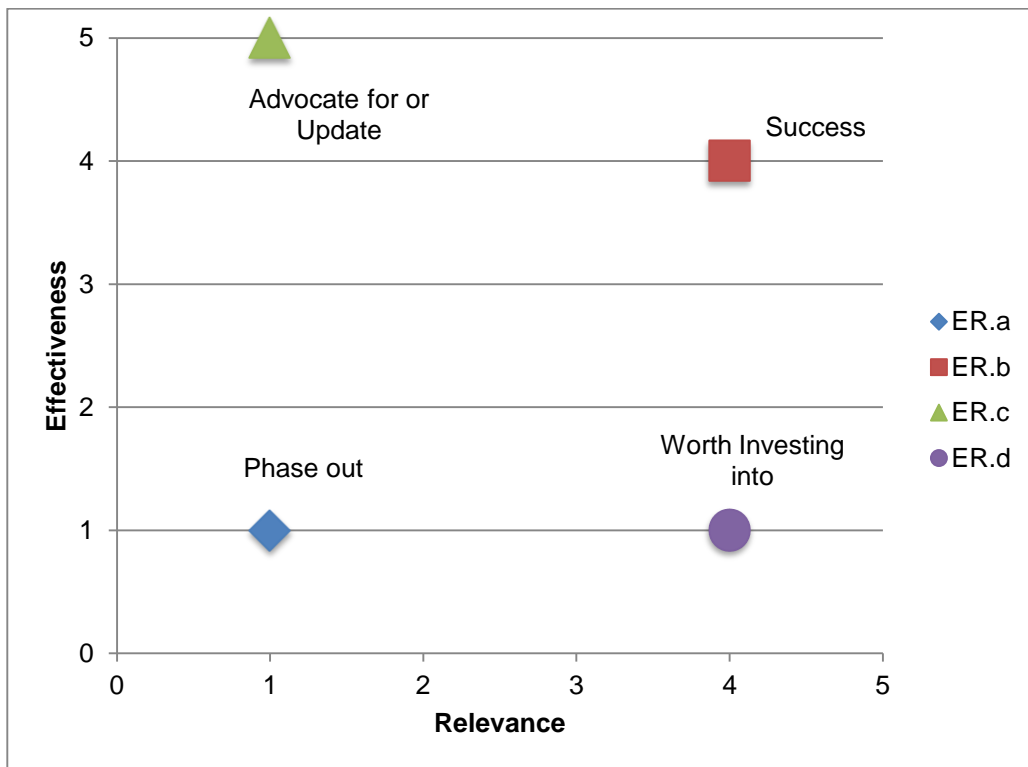
1. Applying the principles of multi-criteria decision analysis

As illustrated in this synthetic analysis, rating interventions based on clear criteria can go a long way in providing important information to decision makers on the worth and potential of interventions. The logical next step would be to use the synthesized information, and in particularly the ratings, to compare programmatic areas of work. Such comparisons can helpfully support decision-making processes on strategic priorities and the strategic allocation of resources. The literature provides multiple tools, decision rules and algorithms on how to deal with multiple criteria, the relative weights attributed to criteria and, ultimately, arriving at an overall ranking of programme areas (based on the criteria and weights). A very simple illustration of a decision support framework with simple decision rules for the case of comparative analysis of programme areas based on two (main) criteria only (i.e. relevance and effectiveness, is the following. For example, to support strategic decisions it may be useful to classify programme areas into four quadrants, based on their ratings (see Figure 30):

- ERs with low relevance and low effectiveness: where withdrawal should be considered;
- ERs with low relevance and high effectiveness: which are in need to be reformulated or repositioned;

- ERs with high relevance and low effectiveness: where further investments and capacity enhancements should be considered;
- ERs with high relevance and high effectiveness: which should be maintained and branded as the core of the Organization's work.

Figure 30. Two by two decision matrix with fictitious Expected Results



2. Complementary roles between evaluation and other information channels to support comparative programme assessment in the framework of the SRR

Ideally, more than two main dimensions should be included in assessing and comparing programmatic areas of work. In the framework of the Strategic Results Report (SRR), UNESCO is in the process of adopting and applying a list of criteria for introducing new programmes and deciding on whether to maintain or discontinue existing programmes (197 EX/5 Part IV, p. 29). The document lays out five main criteria, as well as a number of sub-criteria against which programmatic areas should be assessed. Furthermore, several information channels are listed to provide the evidence and justification behind each rated criterion, including Programme Implementation Reports (PIRs), evaluations, audits, programme reviews and stakeholder surveys. Table 12 lists the criteria retained for the SRR.

As demonstrated in this study, evaluations have a comparative advantage in providing evidence on a number of dimensions, most notably "relevance", "demonstrable contribution and tangible results", and "niche/added value". In addition, evaluations can provide substantive complementary evidence on a range of other criteria, including on capacity to deliver, and some aspects of sustainability. Table 12 lists the possible role of evaluative evidence in programme assessments on the various criteria.

In a decision-making process, more importance could be given to specific criteria in comparison to others (this can technically be achieved through weighting techniques). For example, a particularly

important criterion is UNESCO's niche/value added, which refers to work that UNESCO is ideally positioned to carry out in comparison to other institutions.

Evaluation reports can bring to the table a nuanced and empirically substantiated perspective on UNESCO's (potential) niche in terms of, for example:

- Uniqueness of UNESCO mandate (e.g., only UN agency with a clear mandate on CLT, or communication);
- Uniqueness of UNESCO's lead role (e.g., lead agency for Culture and Development Window of the MDG-F, lead agency for EFA);
- Uniqueness of UNESCO's networks (e.g., UNESCO has a unique relationship with a particular ministry and presence in a given community);
- What others are not doing (e.g., no other agency was intervening in region x on topic y); and
- Uniqueness of UNESCO's programming approach (UNESCO's intersectoral, interdisciplinary, and participatory programming).

Table 12. Potential contribution of evaluative evidence to comparative programme assessment

Key criteria proposed in the 197 EX/5 Part IV:	Evaluation as a primary information channel	Evaluation as a complementary information channel
1. Relevance:		
• Alignment with Major Programmes' objectives	X	
• Contribution to the achievement of SDGs and/or national priorities	X	
• Focus on global priority Africa and Gender Equality	X	
• Focus on target country/population groups (LDCs, SIDS, Youth, etc.)	X	
2. Capacity to deliver:		
• Staff capacities and expertise		X
• Partnerships		X
• Substantial extrabudgetary resources		
3. Comparative advantage/ added value :		
• Programme/project uniqueness/niche	X	
• Programme/project targets emerging issues in UNESCO's fields of competence	X	
4. Demonstrable contribution and tangible results:		
• Tangible and measurable results, directly attributable to the programme	X	
• Degree of visibility	X	
5. Sustainability:		
• Financial factors: UNESCO's financial capacity to maintain programme delivery and results and beneficiaries' capacity to sustain the programme		
• Partner country ownership: Level of Member State's ownership or participation in the programme	X	
• Exit strategy: existence of a plan for exit/sustainability with well-defined time limits and exit points		X

3| KEY FINDINGS

Finding 6: In contrast to self-evaluation and self-reporting, an external evaluation is an independent inquiry based on a systematic process of data collection and analysis. Consequently, evaluations have the potential to provide more credible evidence on a number of strategic performance issues, including outcome (expected results) achievement. Given that there is currently no evaluation strategy at the ER level, assessment of effectiveness and relevance at the programmatic area level has to be extrapolated from a rather patchy evidence base. Nevertheless, the illustrative assessment of the programmatic areas that meet the minimum requirements for synthetic review demonstrates that it is possible to distinguish underperforming from well-performing ERs.

CHAPTER 5. SYNTHESIS OF CROSSCUTTING ISSUES

To identify systemic crosscutting issues from evaluation reports we look at a sample of 68 evaluation reports (see Chapter 1 for the rationale behind this particular sample). In this chapter, we identify crosscutting challenges and enabling factors that were most frequently identified in the sample of reports under review. As highlighted in Chapter 1, there is a bias in evaluation reports towards providing more thorough and in-depth analyses of factors that were found to hinder implementation processes and processes of change, rather than enabling factors. This bias is unavoidably reflected in this synthesis, with more examples of challenges than enabling factors.

1| CROSSCUTTING CHALLENGES

a. A lack of strategic focus [n = 22]

A recurrent set of findings and recommendations in evaluation reports, most notably corporate evaluations, has to do with UNESCO's difficulty in focusing its work on a limited number of programmatic areas where it can make a difference. This theme is further articulated in two sets of issues: (i) tensions between strategic planning and the need to be opportunistic in resource mobilization; and (ii) the lack of clear strategic directions at various levels of intervention.

- ***Tensions between strategic planning and the need to be opportunistic in resource mobilization***

The reduced Regular Programme budget of the Organization and the increasing need to raise extrabudgetary funds has further decoupled the UNESCO system, increasing the gap between official strategic planning (as embodied by the C/4 and C/5) and the reality of "opportunistic" programming to respond to donors and raise additional donor funding. This intrinsic tension is well-illustrated in the 2015 evaluation of the CapEFA: *"The longer-term CapEFA vision on up-stream capacity development processes is difficult to operationalise when financial commitments to the programme are insecure and often tied to annual or biannual commitments pledges."*

The same tensions between establishing and pursuing a clear strategic focus and responding to the diverse and sometimes divergent needs of donors exist for UNESCO Category I Institutes. For instance, the evaluation of UIL concluded that the Institute's mandate requires better clarification. It also noted that the increased dependency on extrabudgetary resources, and consequently on donor-driven priority frameworks may hinder the process of establishing the Institute's strategic focus. The evaluation concludes: *"UIL currently seeks to define its strategic focus and it will be crucial to find the right balance between the new areas of focus, existing expertise and funding opportunities."*

Moreover, several large UNESCO programmes, including Intergovernmental programmes have a distinct operating model, which consists of using seed money to leverage additional funds. This mechanism also amplifies planning challenges and complicates the formulation of clear operational objectives and strategies to achieve them. For example, the evaluation of IHP-VII underscores *"A programme like IHP that is targeting seed money to leverage a greater consortium of partnerships generates significant intangible benefits, which is something to be positively assessed. Yet, the lack of specificity of objectives is definitely an obstacle for the effectiveness and efficiency of IHP activities."*

- **Weak strategic direction in a range of activities**

Issues with weak or absent strategic directions are frequently cited in evaluation reports, indiscriminate of the type of intervention. The word "ad hoc" appears recurrently in the reports to characterize UNESCO's activities: at the level of Trust Funds, intergovernmental programmes, and activities within programmes. The strategic dimensions that were found to be most often missing relate to sustainability strategy (articulating how UNESCO achievements will be sustained over the long term); exit strategy (no clear plan for a sunset clause); capacity development strategy (often a key element in "upstream" interventions); and institutional strategy (e.g., linking work of UNESCO HQ and Institutes, or linking work of UNESCO HQ with Field Offices).

The diversity of activities undertaken within a given programme or under the umbrella of a Trust Fund can quickly turn from asset to liability when the strategic direction is weak or hard to communicate. Weak strategic orientation can have consequences at several levels. First, evaluation reports note that scattered implementation of distinct activities diminishes the potential for impact by spreading capacities and funding too thinly or by missing out on potential synergies between activities. Without a clear strategic orientation, it is also difficult to sequence and connect projects for greater effectiveness. The evaluation of the Japanese Funds in Trust highlights: *"Thin spreading of project funding on many small activities may result in the loss of the overall impact and visibility of the programme."* Another illustration of weak strategic planning comes out of the evaluation of the UNESCO-implemented component of the project "support the media in Iraq." The evaluators noted: *"Good project design needs both horizontal and vertical elements. As detailed in the report, UNESCO's components of the project were implemented in isolation of one another, and in isolation of UNDP components."*

Second, weak strategic direction can also give potential partners and funders the impression of dispersion. For example, in connection to the activities undertaken by UNESCO under the Trust Fund for tsunami preparedness (ESCAP), the evaluation report underscored: *"Currently, the Fund is too diverse to be attractive to the donors. The 16 projects appear more as a collection of excellent activities that are not always planned together in a holistic manner, or for greater impact."* The lack of UNESCO focus and clear strategic vision is also an impediment to effective collaboration within the framework of "Delivering as One". The recent review of UNESCO's field reform in Africa made the following observation: *"Field and HQ focal points interviewed spoke about the conflicts they face when having to state one or two priorities UNESCO would promote for UNDAF developments – would it be Culture or Education or Science? Even when it may seem obvious, the presence of other UN agencies working in the same area with significantly more resources (human and capital) is seen as affecting UNESCO's raison-d'être. Without specific focus, or more representation, UNESCO is missing out on highlighting and driving towards achieving unified results in the sectors in which it operates."*

Weak strategic planning can be connected to another set of challenges which has to do with the governance systems underpinning UNESCO's interventions, some of which are particularly complex. Shortcomings in institutionalized consultation, divergence of opinions between members of steering committees or intergovernmental bodies, can result in weak strategic development. For example, the evaluation of the African World Heritage Fund emphasizes: *"While the coordination and interaction at operational level is excellent, the Fund and UNESCO both highlighted shortcomings with respect to strategy development and joint programming."*

b. Limited financial and human resources [n = 23]

A key challenge hampering UNESCO's capacity to make a difference is the Organization's limited human and financial capacities. Expressions such as "the Organization is spread too thinly", "vacancies

in key positions were not filled", and "the project faced limited budget and human capacity" are recurrent across sectors and areas of work. Limited staff capacity and financial resources are considered particularly damaging when UNESCO intervenes in one of its niche areas, or in an area in which it has acquired a high profile and reputation. For example, in the evaluation of the project "strengthening capacity to combat drought and famine in the Horn of Africa", the SC sector's intervention was praised by many stakeholders but the evaluators also highlighted that *"the project faced limited budget and technology and human capacity to meet the very expensive costs and technical complexity of drilling deep boreholes in the remote rural areas."*

Maintaining or achieving the status of "center of excellence" or of "leading agency" in a particular domain requires a critical mass of expertise, the right skill mix and the necessary human and financial capacities to respond to stakeholder demands. In several of UNESCO's flagship programmes, evaluations underscored that senior expertise is insufficient and key gaps in the skill mix are apparent, thereby damaging the credibility and reputation of UNESCO. For example, the 2012 IHP-VI evaluation noted that *"the current staff numbers, their specialties and profiles at both the IHP-Secretariat and regional level appear to be inadequate to effectively respond to the challenges resulting from IHP successes and growth"*. In turn the corporate evaluation of IHP Phase VII emphasizes: *"Underfunding, staff reductions and the recent lack of a permanent Director have all contributed to management problems at the Secretariat. Staff workloads are high, evaluation procedures ineffective and bureaucracy so heavy it is detrimental as well as not delivering the intended benefits."*

Moreover, what emerges from the series of evaluation reports on the ED sector's Category I institutes is that the latter are not spared from the issue of limited capacity. This conundrum is well articulated in the evaluation of UIL: *"[The institute's] senior expertise is thinly spread in most areas and it still has to find ways to scale up and systematically develop a critical mass of capacities at national and regional levels in a more comprehensive manner."*

Within UNESCO, Field Offices are where human and financial capacities are most often regarded as weak. UNESCO's field presence, in particular in Africa is seen as particularly limited, especially compared to other UN agencies. The Independent External Evaluation of UNESCO, among other reports, noted that it was negatively affecting the organization's reputation and effectiveness. The weak field presence is seen as a particular problem when UNESCO engages in "downstream" operational activities. For example, one evaluation in the CLT sector underlined that *"UNESCO's capacity to support practical cultural heritage protection activities, presents more challenges than the management of the heritage conventions."* The constraints in human and financial resources was apparent in some evaluations of joint programmes as well, where UNESCO is described as *"unable to take a leading role towards particular outcomes of the UNDAF frameworks because of its lack of funding and technical capacity."*

c. Issues of coordination and strategic alignment [n = 17]

One of UNESCO's five main functions is to act as a convening power for international action. Apart from its Secretariat and Field Offices, the wider UNESCO network is made up of a range of entities, spanning Institutes, National Commissions, Chairs, and an array of other partners involved in its numerous intergovernmental programmes. This large UNESCO universe casts a wide net of actors in the different programmatic domains within the mandate of the Organization. In addition, UNESCO increasingly works in partnerships including with other UN organizations. While working through this very wide and diverse network of partners is an undeniable strength of the Organization, it also poses particular challenges in terms of coordination, synergies and joint planning. A number of evaluation reports highlighted these challenges, starting with intersectoral collaboration across the five sectors within the UNESCO Secretariat.

- ***Intersectoral collaboration***

A number of corporate evaluation reports have highlighted the difficulty for UNESCO of working across sectors. The organization's structural set-up in five different sectors at Headquarters, each one with its own budget, reporting lines, priorities, partners, make it very difficult to work intersectorally and transversally. Conversely, evaluations note that there is more ease and potential for effective intersectoral work within Field Offices. The limited number of staff, the common reporting lines to Field Office directors, more flexible processes and the more direct work with partner governments lend themselves to enhanced collaboration across sectors.

A number of other structural factors come out of evaluation reports as hindering intersectoral work. First, there is currently no platform to formalize, facilitate and incentivize intersectoral work. During the 2008-09 biennium, intersectoral platforms (ISPs) were established as a response to the evident need for UNESCO sectors to collaborate more explicitly across sectors to achieve the overarching goals of the Organization. The ISPs have since been discontinued and have not yet been replaced by another mechanism. Both evaluations of Priority Africa and Priority Gender Equality conclude that the potential of these priorities to become a platform for intersectoral programming through their "special programmes" did not materialize.

Second, there is an absence of incentives to work intersectorally, couched in the vertical sector silos, which are not counterbalanced by horizontal incentives and structures to work across sectoral boundaries. There is no dedicated funding mechanism to support intersectoral work, despite additional transaction costs that are attached to working beyond the natural boundaries of a sector, in terms of time, social capital, and administrative complications. Moreover, evaluations have pointed out that staff contributions to intersectoral work are not part of job descriptions, with few incentives for staff, particularly from smaller sectors to engage in intersectoral work, as they often do not get the "credit" for the results achieved.

This combination of challenges is well summed up in the evaluation of "Culture for Sustainable Development": *"Work on culture and sustainable development is, by definition, inter-sectoral. This seems to be obvious, and yet in reality inter-sectoral (or even trans-sectoral) cooperation between culture and other sectors is rare, both in policy and in implementation. Working inter-sectorally within UNESCO has always been a challenge, and while several attempts have been made in the past to find solutions to this problem (inter-sectoral platforms etc.), examples of successful sustained inter-sectoral work that go beyond cooperation in the context of an event or publication are still rare. "*It is also echoed in the evaluation of UNESCO's Strategic Programme Objectives 12 and 13: *"The organisational structure and 'culture' of UNESCO is also considered to be a constraining factor, both at HQ and in the field. Sectors tend to work in silos, though CI could be said to be the least silo-like of all the sectors, with its background and legacy based on information access and provision to meet widely differing sectoral, public and social goals. In the field, however, formal lines of communication are between specialist sector field officers and HQ sector staff and rarely across sectors within a field or cluster office. Budgets for intersectoral work are not available so efforts to identify resources for any collaborative opportunities can face significant procedural hurdles."*

- ***Cooperation with the field network***

Issues of coordination between Headquarters and Field Offices were addressed in a number of evaluations, most notably with regard to: the existing mismatch in the nature of activities carried out in the field and at HQ; and the lack of clarity and complexity of accountability lines. While on the one hand, HQ tends to engage in political and policy-oriented work, Field Offices are in charge of

programme delivery in the field. Yet, they often also constitute the first point of contact with national governments. Consequently, Field Offices are *de facto* often engaged in a host of activities, and as a result spread very thinly. The review of Field Reform in Africa pointed to the dispersion of efforts of Field Offices, and to the need for a clear plan to limit the engagement of Field Offices to fewer strategic priorities, that would make optimal use of existing human capacities, found not only in the Field Offices but also in UNESCO's wider network (such as National Commissions, UNESCO Chairs, and UNESCO Category II Centers).

Weak coordination and support from HQ to the field network also affect the implementation of large and complex intergovernmental programmes that are implemented at various levels (global, regional and national), such as IHP. This type of programme requires close interactions and collaboration among the key players at different levels. The evaluation of IHP-VI notes: *"There is no agreed framework for the IHP Secretariat staff to provide technical back-up support and supervision to regional IHP staff to ensure that IHP activities are effectively and efficiently implemented at the regional and national levels. [...]The objectives of the IHP VI could have been better achieved with a better coordination of the UNESCO regions and sub-regions."*

In addition, the coordination and collaboration between UNESCO Category I Institutes and Field Offices is uneven. While some Institutes work in conjunction with Field Offices on a regular basis, others rarely partner with them. On the other hand, the existing contribution of Category I Institutes to field work is not optimally exploited by the Organization, as highlighted by the recent review of the Field Reform in Africa: *"The role, potential and contribution of Category I Institutes, especially in the education area, has not been fully exploited. A number of Category I Institutes are engaged in various activities in Africa, most visibly IICBA in Addis Ababa, ranging from capacity building (e.g. IIEP offered a large volume of customized training in educational planning in Benin, Rwanda and Namibia), policy development (UIL contributed to the Education Sector Analysis and the preparation of an Education Sector Development Plan for Tanzania) and the production of education-related statistics (UIS provided onsite support to 18 countries across the region on UIS survey instruments, data quality assessments and out-of-school children methodology). The review found very little recognition of the role that Category I Institutes could play in the reform of the field network."*

Yet, there are clear possibilities for further involvement of Institutes in the work of Field Offices. For example, the evaluation of IIEP concluded that, *"[t]here is a potential for improving the field offices' involvement in the implementation, support and follow-up of IIEP's activities. Given the fact that educational planning constitutes one of the core areas of UNESCO's upstream policy work, capacities within UNESCO Field Offices could be strengthened with the support of IIEP for field offices to better play the role of antennas and co-implementers of UNESCO's mandate in educational planning."*

The collaboration between the UNESCO Secretariat and National Commissions was the object of a review in 2011. The review highlighted many instances of effective cooperation, yet it concluded that the Secretariat does not use the network of National Commissions to its full potential. To increase the effectiveness of the cooperation, a number of challenges need to be overcome including: the lack of clarity with regard to the roles and responsibilities of National Commissions; and effective organization-wide working processes, especially in terms of information sharing and knowledge management. The review also highlighted the need to address a number of strategic considerations, including a shared vision and understanding of the objectives of the National Commission Networks, the functions of its members, and the multi-polar collaborations, not only between National Commissions and the Secretariat, but also among National Commissions and with other parts of the wider UNESCO networks.

- ***Working within UNESCO "families"***

Whether it is the UNESCO "Water family", "Culture family", or "Education family", the collaboration within UNESCO's large networks (including HQ, Field Offices, intergovernmental bodies, Category I institutes and Category II centers, National commissions, UNESCO Chairs, expert committees, etc.) operating in a given programmatic area, is systematically characterized as "sub-optimal" or "dysfunctional". Potential synergies are not fully exploited, and evidence of successful cooperation often points at individual initiatives on the side of either UNESCO staff or Chairs and National Commissions, rather than a more concerted effort to engage the various members of a given UNESCO "family". Yet, as the IEE already highlighted, and as echoed in many evaluation reports reviewed for this exercise, for UNESCO to increase its effectiveness it needs to exploit all of its existing assets.

One of the most recurrent observations stemming from the review of the Education Sector Category I Institutes concerns the lack of clarity on how the Institutes are expected to contribute to a given UNESCO sector's objectives. How Institutes' strategies "relate to", "feed into", "subscribe to" UNESCO's overarching sector strategies is often characterized as confusing or unclear. In addition to issues with the alignment of objectives, the reports also identify a number of challenges to the effective collaboration between entities of the UNESCO "Education Family". In Box 9 we cite a few examples stemming from three distinct reviews of Category I Institutes.

Box 9. Examples of challenges of working within UNESCO "families" stemming from Category I institutes evaluations

- **Evaluation of UIL**

"Exactly what the Institute is expected to contribute to the ED sector's overall objectives remains more vague. For the ED Sector's new 2020 strategy, which, among others, is expected to highlight 'fostering lifelong learning opportunities for all', it is therefore of utmost importance to strategically position the Institute and to clearly define mutual expectations in a complementary manner... Despite HQ's appreciation of UIL's work, the visibility and use of UIL's capacities by HQ continues to be limited. In a number of areas, most notably fundraising, competition rather than collaboration appear to be the practice."

- **Evaluation of IIEP**

"Notwithstanding the high level of appreciation within HQ for IIEP's work, the visibility and the use by HQ has so far been rather limited. Moreover, despite recent improvements in the interaction between HQ and IIEP and attempts to better clarify their respective roles, there is still overlap/partly competition as well as a lack of clarity on the division of labour between IIEP, HQ and UNESCO Field Offices. In addition, the relationship between IIEP and HQ has been characterized by a degree of competition in a number of areas, most notably fundraising."

- **Review of UNESCO-IHE**

"UNESCO-IHE's work is highly relevant to the mandate and objectives of the International Hydrological Programme. However, the work of the former is not mentioned in much detail in the IHP-VIII (draft) Strategic Plan and there is inter alia no reference to the Global Campus initiative. This is rather remarkable given the fact that UNESCO-IHE as a Global Campus is premised on the idea of enhancing the Institute's reach in the various UNESCO regions, hence considerably modifying the terms of its affiliation to the 'UNESCO Water Family'. From the Institute's perspective, with respect to the main strategic directions that the Institute is foreseeing for the next few years, references to UNESCO partners (i.e. notably the IHP Secretariat) are conspicuously absent from the Institute's Strategic Plan for 2020."

The need, yet also the complexity, of working effectively within the UNESCO "family" is nowhere more explicit than in the work of intergovernmental programmes, as illustrated in Box 10. Indeed, intergovernmental programmes such as IHP rely on national committees and other entities to

implement their activities. For example, the evaluation of IHP-VI highlighted that proper coordination with National Commissions on the one hand and water ministries was a key ingredient to effective delivery: *"At a regional level, linkage and collaboration between UNESCO National Commissions and water ministries varies greatly from country to country and this had a big bearing on the performance of IHP national committees and hence implementation of IHP activities"*. Similar issues were highlighted in another largely decentralized intergovernmental programme – Man and the Biosphere (MAB). A key finding from the 2014 corporate evaluation of MAB was: *"A significant proportion of biosphere reserves and MAB national committees are "disconnected" from the World network of Biosphere Reserves"*.

Box 10. Examples of challenges of working within UNESCO "families" stemming from IHP-VII evaluation

After the global network of IHP rapidly expanded over the past decade, during which two thirds of all water centres and chairs were added, it became increasingly difficult and burdensome to maintain coherence and meaningful collaboration between members of the large and diverse network. Special and continuing — as opposed to ad hoc — attention is required to reinvigorate links between all levels of the global network and create the synergy needed to halt further disintegration. Gaps are already apparent in communication: lack of knowing who to contact, lack of appreciative feedback on jobs well done, lack of participation in activities, groups going their own way or transferring allegiance to other organisations, and a commonly expressed feeling that politics are often hampering the development and application of the water science. Water-related Centres and Chairs can play an important role in IHP implementation and constitute a key strength of the UNESCO network. However, the unchecked proliferation of UNESCO-affiliated Centres and Chairs carries a reputational risk for UNESCO, especially in cases where new additions deviate from common goals or are inactive altogether.

The cooperation with UNESCO Category II centers is portrayed as even more tenuous. While Category II centers' strategies are supposed to be closely aligned with UNESCO's areas of priority, this is often not the case and the resources and expertise of Category II centers are often not exploited to their fullest potential. The review of the Institute for African Culture and International Understanding (IACIU) pointed to the following issue: *"Since its inception, therefore, the Institute has effectively not received, or indeed requested any operational input from UNESCO HQ to coordinate programmes, or any technical advice or guidance on how it might best have contributed to UNESCO's programme priorities and expected results. One consequence of this has been that the IACIU international forums and meetings on the 2005 Convention (2012 and 2013) were organised without consultation with the Secretariat of the 2005 Convention in UNESCO HQ."*

- **Working with other UN agencies**

The IEE noted that UNESCO was leveraging the opportunities offered by the UN reform process and was seen by several UN partners as playing an active role in the "Delivering as One" process. This active role opened new funding possibilities for UNESCO, intensified joint work with other UN agencies, particularly through the MDG-F Culture and Development window. The opportunity to collaborate with other agencies showcased the capacities of the strongest members of the broader UNESCO network, most notably some Category I institute, and programmes supported by intergovernmental bodies. Yet, a number of coordination issues came to dampen some opportunities of working effectively with other agencies.

For instance, evaluations that took place in the framework of the MDG-F found a deficit in strategies and methods for effective coordinated work that would have guaranteed better interagency synergies. As a result, a number of joint interventions did not appropriately build on the comparative advantages of each agency. For example, the evaluation of the MDG-F Joint Programme "Creating

opportunities for youth employment in South Sudan” highlights: *“The roles of UN agencies did not reflect their comparative advantages. For example, ILO would have been better placed to handle Labour Market Surveys instead of UNDP; and UNICEF was allocated funds for livelihood training instead of ILO, UNIDO or UNESCO. In addition, given that ILO and UNIDO had similar areas of work in the South and North respectively, it would have been more efficient for them to swap activities when the programme was divided into two.”*

Similar coordination issues were apparent in a number of joint operations taking place outside the MDG-F framework. For example, the evaluation of the UNESCO-implemented component of the project “Support to the media in its role of fostering peace and democracy in Iraq” highlighted *“Without a clear articulation of responsibilities and accountabilities, collaboration between agencies is difficult to establish. The available evidence suggests that there was very little coordination or collaboration between UNESCO and UNDP on this project and despite relatively thorough descriptions of their roles in the project document, the actual agreement between the two agencies is just two pages plus budget, none of which details operational guidelines for how the two agencies might work together.”*

d. Dispersed governance systems [n = 14]

Another set of crosscutting challenges recurrently mentioned in evaluations fit under the broad umbrella of “governance”. These challenges relate to:

- Overlapping mandates
- Confusion about who is in charge of steering a programme or institution
- Administrative supervision or tutelage issues
- Insufficient oversight of Category II Centers

Governance issues are most often cited in evaluations of intergovernmental programmes (e.g., IHP, MAB, Culture Conventions, etc.). These programmes tend to encompass multiple levels of governance, endowed with several advisory bodies, steering committees, and reporting mechanisms, often resulting in confusing decision-making processes and diffuse accountability. In addition, and as mentioned above, these multi-layered governance structures can actually stand in the way of achieving a more strategic focus in the large and diverse programmes. These issues are well-captured in the Evaluation of Culture for Sustainable Development: *“the structural make-up of the standard-setting work along Conventions— with each Convention having its own governing systems, advisory bodies, constituency, reporting system, etc.—makes it difficult to work across Conventions, and sometimes even creates a disconnect within conventions.”*

The representativeness and perceived fairness of governance structures is also of primary importance to the effective participation of member countries, and ultimately has some bearing on the impact of programmes. For example, the evaluation of IHP-VI notes that membership in the IHP Council and/or Bureau is a key mechanism to secure country ownership of IHP activities. Yet expanding these two governing bodies too widely, can also result in enhanced confusion and dilution of decision-making power: *“The major issue affecting IHP that was raised by almost all the people and organizations is its inadequate governance mechanism. Current governance structures do not enable Member States to effectively participate in how the programme is planned or implemented. In most regions, countries would actively participate in IHP activities at the global level if they were members of the IHP Council and/or Bureau but will not otherwise, and that members to the IHP Council and Bureau normally represent their countries’ interests rather than regional interests due to lack of formal regional and subregional governance structures. This matter is one of constant debate.”*

If representation in steering bodies is perceived as unfair by participating Member States, this can even result in their disengagement from the programme, as noted in a more recent evaluation of IHP-VII, in the following terms: *"Many Member States amongst the LDCs are not engaging because the current hierarchically organised decision-making system is perceived as lacking representation. Much of sub-Saharan Africa, with the notable exception of Kenya, has virtually detached itself from IHP and is increasingly relying on its own structures, like the African Ministers' Commission on Water (AMCOW)."*

On the other hand, in some intergovernmental programmes with a decentralized decision-making process, such as CapEFA, capacity and leadership issues can also arise, negatively affecting the steering of programmes: *"Although CapEFA interventions are derived from national country policies and planning and UNESCO's approach promotes national ownership, in practice leadership and pro-activeness to make informed choices on CapEFA differ per country and is an issue of concern."*

Governance issues were also prominent in the evaluations of Category II centers. A number of reports highlighted the limited oversight from the UNESCO Secretariat on Category II centers. For example, the evaluation of IHP-VII highlights: *"Around half of all Category II Centres are inoperative and the written evidence of contributions to IHP corroborates this state of affairs. Stricter monitoring is needed and non-functioning Centres should have their UNESCO accreditation withdrawn."* Similar issues are echoed in the evaluation of the Category II center IACIA: *"At present there is little distinction between management and governance in the IACIU; the Executive Board appears to duplicate much of the governance function and meets only once a year, which is not helpful for an executive or management body. The IACIU is too small to need both a Governing Board and an Executive Board and would benefit from a more flexible, smaller Management Committee that would meet more often and support the Director in implementation of the policies and programmes approved by the Governing Board."*

The somewhat unproductive competition for the oversight and governance of UNESCO-affiliated entities was also highlighted in the evaluation of the Pole de Dakar, with two unresolved visions of the governance structure of the Pole. On the one hand, one vision was to merely associate the Pole de Dakar to the Regional Bureau for Education (BREDA) also based in Dakar, with the Pole preserving its autonomy in terms of objectives, strategy and funding. On the other hand, a second vision was to fully integrate the Pole into the BREDA. The impossibility to solve this tension led to confusion and a certain level of acrimony.

Governance concerns also surfaced in evaluations of partnership work and joint programmes. For example, the partnership agreement between the WHC and Panasonic was deemed particularly weak and lacking a formal structure, mandate and procedural framework, which resulted in confusing planning and a lack of clarity on roles and responsibilities. Similarly, confusion about the role of the "lead agency" in the Culture and Development Window of MDG-F was also often cited in evaluation reports. For example, a number of evaluations noted the mismatch between the guidelines about the role of the lead agency in Joint Programmes, and the lessons learned from experience of MDG-F operations. While the former specifies that the role of the lead agency is to lead the collaboration between the different partners, good practice standards emphasize the importance for the lead agency not to "manage" the intervention.

e. Operational challenges [n = 20]

In addition to strategic and governance issues, a number of more practical and operational challenges also surfaced from the evaluation reports. These operational challenges appear at various stages of the project cycle and span a range of operational aspects, from length of engagement, to ensuring timely and proper disbursement of funds, to contractual and hiring impediments (to name a few of the most recurrent issues). The reports often attribute these difficulties to "UNESCO's internal processes", without going into more details. Expressions such as "heavy bureaucracy", and "inefficient and cumbersome processes", are often mentioned.

For instance, multiple reports diagnosed that the centralization of staff in headquarters and the heavy administrative and procurement systems can lead to the slow disbursement of funds. To take only one example, the evaluation of the FETWater programme found that *"operations have suffered from delays in the release of funds and consequently the implementation of the projects of the different networks were affected. This uncertainty in the flow of funds had adverse effect on the ability of FETWater networks to deliver on time."*

Another frequent operational challenge cited in evaluations is the absence of a proper exit or "hand over" mechanism to ensure the continuity of engagement. A related issue has to do with the length of engagement in activities. UNESCO's involvement was often deemed to be too short, especially in the case of Joint Programmes. Taken together, these various issues can be summed up in an overarching problem of the organization, which was well-documented in the Independent External Evaluation, namely, the absence of a proper project cycle, with preparatory work, a clear logic of intervention and implementation plan combined with a proper monitoring and evaluation (M&E) plan and sunset clauses. Weak M&E was frequently highlighted, with the most often cited characteristics of weak M&E being: (i) the absence of a "theory of change" or proper "intervention logic"; (ii) the absence of an indicator framework or monitoring procedures to measure change over time; and (iii) weak systems to track progress, identify challenges and ensure course corrections. The two following citations exemplify these shortcomings:

"The project theory does not explicitly explain in what way the project's output objectives will contribute to the outcome objective. It is therefore very difficult for the project implementation team to monitor results and achievements regarding the outcome objective and bridging objectives." (Evaluation of UNESCO-implemented component of "support to media in Iraq")

"There appear to have been limited mechanisms for performance monitoring to ensure that individual themes deliver their expected outputs, that staff at various levels performs to their expectations, and that the programme performance is monitored by UNESCO governance structures. It was reported that performance indicators are mainly tagged to the level of utilisation of funds, which does not necessarily indicate that the expected outputs were delivered and in a cost-effective as well as efficient manner." (IHP-VI external evaluation)

f. Challenges in the implementation of Priority Africa

Two corporate evaluations focused on UNESCO's Priority Africa and on the reform of the Field Network in the region that was intended to support the priority. While both exercises conclude that a prioritization of UNESCO's resources and capacities towards the region continues to be warranted, they also both highlight a number of structural challenges that stand in the way of an effective prioritization. Most notably, the Evaluation of Priority Africa found that *"the mechanisms which are meant to provide impulse and substance to Priority Africa have had very limited success."* Priority Africa has not led to a substantial increase in the decentralization of human and financial resources to the region, nor has it led to substantive improvement in results achieved. While most of the challenges identified in the evaluation, are not unique to the implementation of Priority Africa— such as the limited decentralization of authority and financial resources to Field Offices, a lack of

programmatic focus and inefficient administrative procedures— these systemic challenges were found to particularly affect the work of the Organization in Africa. In addition, the roles and responsibilities within the framework of Priority Africa, between the various entities in charge of steering, fundraising, monitoring and reporting, coordinating partner actions, and implementing programmes, is confusing.

The reform of the Field Network in the region was expected to address some of the key challenges identified in the evaluation of Priority Africa, with a particular emphasis on delegating responsibilities and ensuring appropriate staffing in all strategic and operational functions (e.g., implementation, fundraising, partnership building, and M&E). However, the 2015 review of the UNESCO's Field Reform in Africa found that many of the promises of the reform did not materialize due to important hurdles along the way. While the basic building blocks of the reform were achieved, including the establishment of five Multi-Sectoral Regional Offices (each staffed with expertise on the five UNESCO sectors and accompanied with an increase in the decentralization of funding to these five regional hubs), a number of challenges were not surmounted and ultimately compromised the success of the reform. The review emphasizes the following institutional constraints:

- The process of reform was significantly delayed, especially in terms of staffing key positions.
- The reform was not complemented by a clear and robust implementation plan and human resource strategy with key objectives and targets.
- There was a discrepancy between the field presence and the coverage of African Regional Economic Communities, or the regional groupings of the UNDG, which hampers well-needed coordination and collaboration with these key partners.
- The field reform has not yet permitted a consolidation of the Field Network with other UNESCO-wide entities based in Africa as a means of bringing a critical mass of expertise in the offices. Most notably, the role and contribution of Category I institutes, especially in the Education area, has not been exploited.
- The managerial responsibility and accountability for carrying out the reform were unclear. In particular, the role of the Africa Department was not spelt out well in the various documents. Oversight was thus diffused and led to inefficiencies.
- The financial support required to fully implement the field reform did not materialize, which translated in a delayed and partial roll-out of the reform, and the envisaged staff movements did not take place as intended.

In addition, from a programmatic perspective, the reform was found defective insofar as it did not help UNESCO sharpen its programmatic focus. The review highlighted the following points:

- The examples of field work guided by explicit programmatic strategies based on country needs were scarce. UNESCO country plans were not systematically up to date and no sub-regional strategies were in place at the time of the evaluation.
- The limited staff capacity in the Multi-Regional Offices and national offices continues to prevent UNESCO from engaging effectively in downstream activities, which remain the types of activities for which extrabudgetary funding is more likely to be available.
- A related issue is that the approaches to extrabudgetary fundraising are often ad hoc and hinders the ability of the Field Offices to meet priority needs of Member States or implement UNESCO's core functions in its areas of specialization.
- There needs to be a continued discussion about whether UNESCO is best fit for purpose at a global and regional level serving as a standard-setting, normative agency whose expertise rests in technical specialization and coordination, or it is best positioned to support mainly national interventions at the request of Member States and/or donors. In this regard, it needs to be recognized that voluntary, mainly bilateral, funding now represents the largest funding

source for activities in the field. Attracting such funding has a risk of UNESCO focusing disproportionately on downstream activities.

g. Challenges in the implementation of Priority Gender Equality

A review of UNESCO's Priority Gender Equality was conducted in 2013 and identified some key challenges that hindered the full realization of the Organization's vision for gender equality, most notably: the strategic framework of the priority and corresponding M&E framework, and the balance and articulation of gender-specific programming and gender mainstreaming. In what follows, we synthesize these key challenges:

- Challenges with UNESCO's strategic framework and M&E for gender equality:
 - Despite the existence of a Gender Equality Action Plan (GEAP), and recurrent mention of the Priority in UNESCO's C/5 and C/4, these three sets of strategic documents are not harmonized and gender is unevenly mainstreamed in UNESCO's sectors.
 - The GEAP does not constitute a useful action plan as it lacks specific targets, timelines, contains too many indicators that are difficult to measure and is too disconnected from the key programmatic documents of UNESCO.
 - Given the lack of integration of the GEAP into UNESCO's C/5 there is also a lack of mechanisms to properly assess progress with respect to Gender Equality.
 - Evaluations need to better integrate Gender Equality in all stages of the evaluation process.

- Challenges relating to gender-specific programming and gender mainstreaming :
 - The principle of gender mainstreaming is not supported by any clear accountability and performance management mechanism.
 - Many staff members do not (yet) have the capacity to properly mainstream gender in their area of expertise. There has been insufficient capacity development for staff.
 - The Gender Focal Point Network is not fully institutionalized, nor does it encompass all the gender expertise within the Organization.

2 | CROSSCUTTING ENABLING FACTORS

A rather clear picture of UNESCO's strengths emerges from the synthesis. What is obvious from the evaluation reports is that one needs to think in terms of a combination of factors that provide the enabling conditions for effective planning, partnership building, programme delivery and monitoring and evaluation. This synthesis echoes the findings of the IEE, in highlighting that UNESCO appears to be strong when:

- It operates in an area where it has a clear mandate to act and is recognized by partners and clients as a legitimate actor.
- It mobilizes a critical mass of staff with adequate experience and expertise.
- It works together across teams and across disciplines within the larger UNESCO family.
- It works well with UN partners and a broader networks of partners.
- It is able to mobilize funds and resources and target them to its priority areas of focus.

In this section we illustrate these enabling factors with evidence stemming from the evaluation reports reviewed for this synthesis.

a. Participatory and interdisciplinary programming practices [n = 13]

One of UNESCO's key assets often recognized in evaluation reports is its distinct programming practices. When done properly, UNESCO is recognized for its distinct interdisciplinary and participatory approach to programming, notably in the design phase. Several reports highlighted extensive consultations with government representatives, scientific institutions, other development agencies and NGOs that were embedded into UNESCO's programmatic approach. Moreover, when UNESCO extends its participatory approach to the implementation phase, by leveraging the experience of local collaborators, it is considered a strong success factor.

A few reports highlighted exemplary preparation processes from a stakeholder participation perspective, which was recognized as a key condition to promote and sustain the broad institutional endorsement and "buy in" to the various projects. For instance, the evaluation of a project aimed at strengthening capacity to combat drought and famine in the Horn of Africa concluded: *"The inclusive approach of the design stage has been continued throughout the project's implementation as the project put much effort into getting many stakeholders involved or sensitized to the implementations of its activities and its achievements. Communication was efficient as a multitude of stakeholders in Nairobi and in the Turkana County was informed both on implementation of activities and their results. Decision-makers at various levels were informed or made aware of the project and the significance of the results."*

The use of participatory methodologies, and the adequate training of staff in implementing them were also praised in the evaluation of the project "Adaptive Learning in Tsunami Preparedness at Community Level In Colombia, Ecuador, Peru and Chile" in the framework of DIPECHO: *"These techniques and the technical capacities of the groups and persons that conducted them has a major positive impact on the specific achievements of the project, and in motivating educational institutions, and the public at large to become involved and to participate actively."*

When Category I institutes manage to integrate a demand-driven element to their principal focus area they are also found to be even more effective. For example, the corporate evaluation of IIEP found that the Institute is particularly demand-oriented in its technical assistance at the country level and when it articulates its training offers. This approach, tailored to actual client needs, was also balanced with the necessity to maintain a strategic focus by setting up clear screening criteria to respond to a myriad of demands.

Finally, when UNESCO prolongs its stakeholders' involvement beyond the design phase, notably by working through local implementing agents to carry out specific activities, this was found to bolster ownership from the targeted communities. When working with local partners with prior experience in the programmatic area and good knowledge of the programmatic context, interventions are found to be more effective at reaching the right beneficiaries and in ensuring the continuity of efforts. For example, the evaluation of UNESCO's work on Culture and Sustainable Development noted: *"UNESCO has also promoted the involvement of local communities via specific projects on the ground, including the support for the participatory elaboration and adoption of management plans in World Heritage sites (e.g. for the Island of Gorée and the Island of Saint Louis in Senegal; and the Historic Centre of Santa Ana de los Ríos de Cuenca in Ecuador, and many others)."*

b. Potential to mobilize and deploy a critical mass of diverse expertise [n = 7]

The reports brought to light a number of instances when UNESCO managed to align funding opportunities with its core programmatic strategy which resulted in innovative and effective operational strategies and bolstered the Organization's impact. For example, the evaluation of Culture and Sustainable Development found that the role of UNESCO as lead agency in the MDG-F window for Culture and Development was a boon for the CLT sector, providing considerable funding to a sector

that is usually underfunded. The CLT sector and its partners took advantage of this funding to engage in innovative initiatives that they could bring to scale. The report also noted that this particular source of funding allowed the sector to work across areas of specialties, combining integrative approaches to tangible and intangible cultural heritage, creative industries, etc. In many ways, the sector instrumentally used the MDG-F to showcase how these various areas relating to Culture and Development could work in synergy and improve effectiveness.

Additionally, when UNESCO deploys a critical mass of diverse expertise and manages to convene the right partners, it is also more likely to be effective. The combination of the right level and type of expertise, including through partnerships, and a sufficient amount of human resources were seen as critical elements to successful delivery and to uphold the credibility of the organization with Member States. The evaluation of ASPnet illustrates this point: *"The evaluation finds that there are at least four major enabling factors determining the positive results of the programme:*

- the presence of a dedicated and capable management structure in Paris*
- the characteristics of the implementing partners*
- a vast and reliable network of actors in the two project sites*
- UNESCO's world-class expertise in developing high quality educational material and in the delivery of effective international and national trainings for different audiences."*

c. Large network of partners [n = 18]

One of the oft-cited strengths of UNESCO is also the breadth and diversity of its networks. When the Organization strategically taps into its broad range of partners and manages to coordinate the contribution of multiple partners well, it has been found to make a significant difference. What follows is a list of cases where strong coordination from UNESCO entities proved fundamental to the effectiveness of a particular intervention:

"Notable successes include the synergy forged between the project's partner institutions and other DIPECHO projects, leading to the consolidation of resources and improvements in tsunami preparation strategies; the cooperation between specialised institutions from each country enabling resources to be pooled and significant advancements to be made." (Evaluation of Adaptive Learning in Tsunami Preparedness at Community Level in Colombia, Ecuador, Peru and Chile)

"During the project design phase all UNESCO field offices were consulted by UNESCO Headquarters and these offices in turn consulted a selection of radio stations for their input. This "North-South" consultation dynamic provided many important and relevant inputs to the project formulation process and helped UNESCO to design activities that were adapted, in general, to the needs of the secondary beneficiaries (with the exception of the online platform, which was a donor request)." (Evaluation of Empowering Local Radios with ICT)

"Successful examples for coordinated delivery are mainly initiated in the field such as a joint initiative involving the expertise of UIL, IIEP, UIS, together with BRENDA, and Pole de Dakar for activities in DRC managed by the Kinshasa office." (Review of UNESCO Institute for Lifelong Learning)

The evaluation of the Japanese Funds in Trust provided another example of a strong and coordinated intervention pulling resources and expertise from various parts of the wider UNESCO network. The following citation illustrates the effective coordination across various entities of UNESCO and with partners: *"Project proposals were prepared by UNESCO Office Jakarta, in its function as the Asia and the Pacific Regional Bureau for Science, and submitted via BSP/CFS to the donor (MEXT, Japan). In the preparation of these proposals, UNESCO Office Jakarta proactively sought inputs from UNESCO field offices in the region, UNESCO/IOC-WESTPAC, beneficiary countries and other partners, including*

Category II Regional Centres, International Council for Science (ICSU), Universities and Research Centres, and Japanese partners."

3 | KEY FINDINGS

Finding 7: Across sectors, levels of interventions and domains of expertise, it is clear that UNESCO has a number of strengths and attractive features, which— when they come together—have enabled important programmatic successes. Chief among these are: UNESCO's participatory and interdisciplinary programming practices, its large network of institutional partners, and its potential to mobilize and deploy a critical mass of diverse expertise.

Finding 8: Nevertheless, the coalescence of these important ingredients for a relevant and effective intervention seems to be the exception, rather than the rule. A number of key structural challenges— indistinctive of sectors or areas of work—have hindered UNESCO's capacity to make a difference. Chief among these are: a lack of strategic focus that affects the quality and potential for impact of its work at all levels of intervention, limited financial and human resources, issues of coordination and strategic alignment that weaken the potential of the wider UNESCO network, dispersed governance systems, and a number of operational challenges.

CHAPTER 6. OVERALL CONCLUSION AND RECOMMENDATIONS

Evaluations are potentially the most credible source of evidence on a number of strategic performance issues, including outcome (expected results) achievement of UNESCO's programmes. However, the current evaluation coverage of UNESCO's programmes is fragmented and uneven. In addition, significant challenges to improving the quality of decentralized evaluations remain. To strengthen the role of evaluations in supporting evidence-informed decision-making, the study recommends the following:

Recommendation 1: The Organization continues to spend too much effort on (micro) activity assessment and reporting at the cost of adequate strategic reflection and assessment at a higher programmatic level. **UNESCO should develop a clear definition of a programme or programmatic area of work** (e.g. around an ER). **Each programmatic area of work should be tied to a standardized information cycle**, constituting the basis for better planning, implementation, monitoring and evaluation, which should include the following elements:

- a strategic analysis, resulting in the development of an intervention logic of the programmatic area of work that clearly articulates the main activities, outputs (deliverables) and outcomes (expected results). This would provide the basis for:
- a results framework with clear and comprehensive indicators at output and outcome levels, which in turn constitutes the basis for:
- continuous monitoring of programme output delivery and outcome achievement, and periodic evaluation of the programme.

The standardization of the unit of analysis (a programme) for planning and information collection purposes, in combination with a harmonized approach to information collection across programmes, closely resembles the project cycle principle that has been successfully adopted in many organizations across the globe. It has the potential to both improve the quality of planning, monitoring and evaluation, as well as lowering the transaction costs for doing so, enhancing UNESCO's potential to become more efficient and effective at the same time.

The framework described above represents a deepening of a process that has already been set in motion through various reform efforts and Executive Board decisions and would strengthen the foundation for supporting the Organization's reform toward improved:

- results-based budgeting;
- evidence-informed decision-making on the strategic allocation of financial resources and the identification of strategic priorities.

Recommendation 2: To improve the role of evaluation to support evidence-informed decision-making, **UNESCO should strengthen its decentralized evaluation system**. More particularly, to improve the quality and coverage of evaluations, the Organization should:

- increase the resources available for evaluation through improved budgeting practices and procedures for extrabudgetary activities;
- strengthen staff capacities for managing decentralized evaluations;
- strengthen the mechanisms for planning, backstopping and information exchange of decentralized evaluations.

Recommendation 3: **UNESCO** (i.e. BSP, IOS and KMI) **should improve the data collection and tracking system of decentralized evaluation reports**. This would allow the Organization to improve its database of decentralized evaluation reports, and consequently the quality and use of periodic meta-evaluations and syntheses of evaluation reports.

ANNEX 1: LIST OF PROGRAMMATIC AREAS

	MLA	Code	Expected Result	Thematic Area
I. Education	MLA1. Supporting Member States to develop education systems to foster high quality and inclusive lifelong learning for all	ED1	National capacities strengthened to develop and implement policies and plans within a lifelong learning framework	Education Policy and planning
			National capacities strengthened to scale up inclusive and gender responsive quality literacy programmes	Literacy
			Capacities of Member States strengthened to design and implement policies aiming at transforming TVET	TVET
			National capacities strengthened to develop evidence-based higher education policies to address the challenges of equity, quality, inclusion, expansion, mobility and accountability	Higher Education
			National capacities strengthened, including through regional cooperation, to develop and implement teacher policies and strategies so as to enhance the quality of education and promote gender equality	Teacher policies and strategies
			Capacities of Member States strengthened to promote, monitor and assess the processes and outcomes of competency-based learning	Promotion, monitoring of competency-based learning
			National capacities strengthened to develop and implement technology policies in education, particularly in teacher training and professional development	ICT and Education
	MLA2. Empowering Learners to be creative and responsible global citizens	ED2	Member States integrate peace and human rights education components in education policies and practices	Peace and human rights education
			Capacities of Member States strengthened to integrate ESD into education and learning, and ESD strengthened in the international policy agenda	Education for Sustainable development
			Member States deliver good quality health education, HIV and comprehensive sexuality education that contribute to healthy lifestyles and gender equality	Health Education
	MLA3. Advancing Education for All (EFA) and shaping the future international education agenda	ED3	Future education agenda and global education policies shaped, drawing on relevant research and foresight studies conducted by UNESCO and other institutions	Foresight and research for EFA
			Implementation of the right to education and progress towards international education goals promoted and monitored, and policy dialogue informed by the evidence generated	Monitoring of EFA
			Political commitment for education reinforced sustained in the global, regional and national development agendas, and cooperation modalities promoted	Cooperation and partnerships for EFA

II. Natural Science	MLA1. Strengthening STI Policies, governance and the Science-policy-society interface	SC1	STI policies and governance bolstered nationally, regionally and globally	STI policies
			Science-policy interface enhanced and sustainability science both promoted and applied	Sciences-Policy interface and Sustainability science
			Mutual engagement of science with society reinforced to promote equity and inclusion of vulnerable groups, including SIDS and indigenous peoples	Vulnerable groups, SIDS and indigenous peoples
	MLA2. Building institutional capacities in science and engineering	SC2	Capacity-building in research and education in the natural sciences enhanced, including through the use of ICTs	Capacity building in research and education in natural sciences
			Interdisciplinary engineering research and education for sustainable development advanced and applied	Engineering research
	MLA3. Promoting knowledge and capacity for protecting and sustainably managing the oceans and coasts	SC3	Scientific understanding of ocean and coastal processes bolstered and used by Member States to improve the management of the human relationship with the ocean	Ocean and coastal research
			Risks and impacts of ocean-related hazards reduced, climate change adaptation and mitigation measures taken, and policies for healthy ocean ecosystems developed and implemented by Member States	Ocean-related hazards and climate change (e.g., tsunamis)
			Member States' institutional capacities reinforced to protect and sustainably manage ocean and coastal resources	Institutional capacity for management of ocean and coastal resources
	MLA4. Fostering international science collaboration for earth systems, biodiversity and disaster risk reduction	SC4	Global cooperation in the ecological and geological sciences expanded	Global cooperation in ecological and geological sciences
			Risk reduction improved, early warning of natural hazards strengthened and disaster preparedness and resilience enhanced	Early warning of natural hazards (e.g., earthquake)
	MLA5. Strengthening the role of ecological sciences and biosphere reserves	SC5	Use of biosphere reserves as learning places for equitable and sustainable development and for climate change mitigation and adaptation strengthened	Biosphere reserves
	MLA6. Strengthening freshwater security	SC6	Responses to local, regional and global water security challenges strengthened	Water security challenges (e.g., drought, flood)
Knowledge, innovation, policies and human and institutional capacities for water security strengthened through improved international cooperation			Capacity for water management	

III. Social and Human Sciences	MLA1. Mobilizing future-oriented research, knowledge and policy-making to support social transformations, social inclusion and intercultural dialogue	SHS1	Future-oriented social science and humanities research on social transformations and intercultural dialogue enhanced through the uses of sustainability science as well as fully inclusive human rights-based and gender-sensitive initiatives to strengthen national social science policy and international scientific cooperation	Social transformation
			Focused initiatives in education, culture, the sciences, communication and information developed that support the emergence of more inclusive societies and greater intercultural dialogue	Intersectoral work for social inclusion
			Capacities of decision-makers, civil society organizations and other key stakeholders strengthened, to design and implement innovative proposals for the development of public policies in favour of social inclusion and intercultural dialogue, particularly targeting disadvantaged populations	Peace and security and intercultural dialogue
	MLA2. Empowering Member States to manage the ethical, legal, environmental and societal implications of scientific and technological challenges with a view to achieving inclusive and sustainable social development	SHS2	Capacities of Member States strengthened to manage bioethical challenges arising from science and technology, operationalize universal bioethical principles, and engage fully in the global bioethical debate	Bioethics
			Ethical, legal and social implications of cutting-edge science, emerging technologies and their applications identified	Ethics
	MLA3. Building policies through a participatory process with stakeholders in both the fields of youth and of sports; supporting youth development and civic engagement and promoting Human-rights	SHS3	Capacities of Member States strengthened to design and implement multi-stakeholder and inclusive public youth policies and young women and men engaged in community building and democratic processes	Youth
			Member States design and implement multi-stakeholder and inclusive public policies in the field of physical education, sports and anti-doping	Sports and physical education
			Human Rights-based approach further integrated in activities across UNESCO's major programmes and in all the phases of programme cycle	Human rights based approach

	based approach in UNESCO's programmes			
IV. Culture	MLA1. Protecting, conserving, promoting and transmitting culture, heritage and history for dialogue and development	CLT1	Tangible heritage identified, protected, monitored and sustainably managed by Member States, in particular through the effective implementation of the 1972 Convention	Tangible heritage: 1972 convention
			Policy dialogue promoted to combat illicit import, export and transfer of ownership of cultural property through enhanced, strengthened and more efficient international cooperation, including the implementation of the 1970 Convention and enhanced capacities of museums	Illicit transfer of cultural property: 1970 convention
			Global, strategic and forward-looking directions developed and applied through the effective implementation of the 1954 Convention and its two Protocols and multiplier effect achieved	Protection of cultural property in the event of conflict: 1954 convention
			Global, strategic and forward-looking directions developed and applied through the effective implementation of the 2001 Convention and multiplier effect achieved	Protection of underwater cultural goods: 2001 convention
			Access to knowledge enhanced through the promotion of shared history and memory for reconciliation and dialogue	History and memory
V. Communication	MLA2. Supporting and promoting the diversity of cultural expressions, the safeguarding of the intangible cultural heritage, and the development of cultural and creative industries	CLT2	National capacities strengthened and utilized to safeguard the intangible cultural heritage, including indigenous and endangered languages, through the effective implementation of the 2003 Convention	Intangible cultural heritage, indigenous and endangered languages: 2003 Conventions
			National capacities strengthened and utilized for the development of policies and measures to promote the diversity of cultural expressions, through the effective implementation of the 2005 Convention	Diversity of cultural expressions: 2005 conventions
V. Communication	MLA1. Promoting an enabling environment for freedom of expression, press freedom, and journalistic safety, facilitating pluralism and participation in media,	CI1	The environment for freedom of expression, press freedom, journalistic safety and self-regulation is strengthened, for both online and offline media platforms, and especially in post-conflict countries and countries in transition, through favourable policies and practices	Freedom of expression, press freedom and journalistic safety
			Pluralistic media institutions are facilitated, including by adoption of gender-sensitive policies and through support for strengthened	Community and pluralistic media and media literacy among youth

	and supporting sustainable and independent media institutions		community media policy and practice, while citizens, and particularly youth, are empowered through enhanced media and information literacy (MIL) competencies	
			Independence and sustainability of national media institutions bolstered, through innovative, policy-relevant, knowledge-enhancing (IPDC) projects and through capacity-building for journalists and journalism schools	IPDC
MLA2. Enabling Universal access and preservation of information and knowledge	CI2		The Open Solutions for Knowledge Societies programme (open education resources, open access, free and open source software, open training platform, open data and Open Cloud) and ICT accessibility, including for the disabled, and for all languages, promoted in Member States	Open Solutions and ICT
			Documentary heritage in all its forms preserved through a strengthened Memory of the World Programme	Memory of the world programme
			Member States supported in implementing the outcomes of the World Summit on the Information Society (WSIS) and universal access to information enhanced, including through the Information for All Programme (IFAP)	World Summit on the Information Society (WSIS) and the Information for All Programme (IFAP)

ANNEX 2: ASSESSING MINIMUM REQUIREMENTS FOR SYNTHETIC ANALYSIS BY EXPECTED RESULTS

Code MLA	Thematic area of ERs	Code	Total	Final Evaluations with quality >4	Final Evaluations with quality > 4 & Independently commissioned	Minimum requirements for synthetic review met
ED1	Education Policy and planning	ED11	15	8	3	Yes
	Literacy	ED12	12	6	4	Yes
	TVET	ED13	9	5	2	No
	Higher Education	ED14	5	4	2	No
	Teacher policies and strategies	ED15	10	5	2	No
	Promotion, monitoring of competency-based learning	ED16	7	4	3	Yes
	ICT and Education	ED17	8	7	3	Yes
ED2	Peace and human rights education	ED28	7	3	1	No
	Education for Sustainable development	ED29	8	4	1	No
	Health Education	ED210	10	4	1	No
ED3	Foresight and research for EFA	ED311	9	8	7	Yes
	Monitoring of EFA	ED312	6	4	3	Yes
	Cooperation and partnerships for EFA	ED313	4	2	1	No
SC1	STI policies	SC11	3	3	2	No
	Sciences-Policy interface and Sustainability science	SC12	1	1	1	No
	Vulnerable groups, SIDS and indigenous peoples	SC13	1	1	1	No
SC2	Capacity building in research and education in natural sciences	SC24	1	1	1	No
	Engineering research	SC25	1	1	1	No
SC3	Ocean and coastal research	SC36	3	1	1	No
	Ocean-related hazards and climate change (e.g., tsunamis)	SC37	8	2	1	No
	Institutional capacity for management of ocean and coastal resources	SC38	3	3	2	No
SC4	Global cooperation in ecological and geological sciences	SC49	5	2	2	No
	Early warning of natural hazards (e.g., earthquake)	SC410	3	2	2	No
SC5	Biosphere reserves	SC511	13	1	1	No
SC6	Water security challenges (e.g., drought, flood)	SC612	7	7	5	Yes
	Capacity for water management	SC613	19	6	6	Yes

(Continued)

Code MLA	Thematic area of ERs	Code	Total	Final Evaluations with quality >4	Final Evaluations with quality > 4 & Independently commissioned	Minimum requirements for synthetic review met
SHS1	Social transformation	SHS11	1	1	1	No
	Intersectoral work for social inclusion	SHS12	2	1	1	No
	Peace and security and intercultural dialogue	SHS13	22	2	2	No
SHS2	Bioethics	SHS24	1	1	1	No
	Ethics	SHS25	1	1	1	No
SHS3	Youth	SHS36	8	3	3	Yes
	Sports and physical education	SHS37	0	0	0	No
	Human rights based approach	SHS38	2	2	2	No
CLT1	Tangible heritage: 1972 convention	CLT11	28	7	5	Yes
	Illicit transfer of cultural property: 1970 convention	CLT12	3	2	2	No
	Protection of cultural property in the event of conflict: 1954 convention	CLT13	3	2	0	No
	Protection of underwater cultural goods: 2001 convention	CLT14	1	0	0	No
	History and memory	CLT15	2	2	2	No
CLT2	Intangible cultural heritage, indigenous and endangered languages: 2003 Conventions	CLT26	22	9	8	Yes
	Diversity of cultural expressions: 2005 conventions	CLT27	34	8	8	Yes
CI1	Freedom of expression, press freedom and journalistic safety	CI11	2	0	0	No
	Community and pluralistic media and media literacy among youth	CI12	3	2	1	No
	IPDC	CI13	5	1	1	No
CI2	Open Solutions and ICT	CI24	3	1	1	No
	Memory of the world programme	CI25	0	0	0	No
	World Summit on the Information Society (WSIS) and the Information for All Programme (IFAP)	CI26	0	0	0	No