



IOM CENTRAL EVALUATION  
July 2025

#### EXTERNAL EVALUATION TEAM

Leticia Bendelac	Team Leader,
Inmaculada Román	Senior evaluators
Carlos Madridejos	Senior evaluators
Lucas Rey	Data Analyst
Salvador Bustamante	Methodological expert

#### IOM CENTRAL EVALUATION DIVISION

Andres Botero	Evaluation Manager
---------------	--------------------

#### EVALUATION REFERENCE GROUP

Angeline Wabanda	Regional Planning, Monitoring and Evaluation Officer
Yves Abdoul Koala	Regional Planning, Monitoring and Evaluation Officer
Theogene Nshimiyimana	Regional Planning, Monitoring and Evaluation Officer
Mariana Sato	Senior Programme Assistaute – IOM Development Fund
Rogers Muti	Monitoring and Evaluation Officer

## **DISCLAIMER**

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the International Organization for Migration (IOM). The designations employed and the presentation of material throughout the publication do not imply expression of any opinion whatsoever on the part of IOM concerning the legal status of any country, territory, city or area, or of its authorities, or concerning its frontiers or boundaries.

## **FOREWORD**

Evaluation contributes to organizational learning while providing decision-makers and stakeholders with independent, evidence-based assessments of IOM's work. To be credible, evaluations must meet UNEG and IOM quality standards.

This second review of IOM evaluations (2020–2024) builds on the criteria developed for the previous meta-evaluation (2017-2019) to:

- assess the quality of evaluations,
- review the relevance and use of the quality assessment tools introduced, and
- examine changes in quality compared to the previous period.

The review finds a modest improvement in the quality of evaluations relative to 2017-2019. However, it highlights persistent weaknesses - particularly the description of the evaluation approach and the comprehensive inclusion of cross-cutting elements - which had already been flagged in the earlier meta-evaluation.

While the introduction of quality control tools has brought greater structure to the evaluation process, the analysis points to the need for more rigorous and standardized quality assurance at key stages - terms of reference, inception reports, and draft final reports. Establishing quality assurance checkpoints is essential to ensure rigour, reliability, and accountability in all evaluations.

The assessment also highlights the need to consistently allocate sufficient time, budget and expertise for evaluations. In particular, it calls for earmarking a fixed percentage of project funds for evaluation, as outlined in the IOM Evaluation Guidelines. Adequate resourcing includes planning evaluations early and ensuring teams have the necessary skills. IOM should offer continuous training programs and mentoring, including for evaluation managers.

I am confident that the lessons and recommendations from this meta-evaluation will help IOM further improve the quality of both centralized and decentralized evaluations, ultimately enhancing their credibility and use.

**Hicham Daoudi**

Chief, Central Evaluation Division

IOM

## **ACKNOWLEDGMENTS**

This meta-evaluation would not have been possible without the invaluable contributions and support from numerous stakeholders. I would like to extend my sincere gratitude to all those who generously contributed their time, expertise, and insights throughout the process.

The evaluation team was led by Leticia Bendelac, Team Leader, and included senior evaluators Inmaculada Román and Carlos Madridejos, data analyst Lucas Rey, and methodological expert Salvador Bustamante, from Artival Research & Evaluation.

I would particularly like to acknowledge the fundamental contributions from IOM staff, including regional planning, monitoring and evaluation officers, evaluation managers and internal evaluators. Special recognition goes to Angeline Wabanda and Theogene Nshimiyimana, Regional Planning, Monitoring and Evaluation Officers, and Mariana Sato, Senior Programme Assistant from the IOM Development Fund. I would also like to recognize Simon Hay, who conducted the editorial review of the report.

The collaborative engagement of all participants was fundamental to the success of this meta-evaluation.

Andres Botero

Evaluation Manager

# CONTENTS

LIST OF FIGURES .....	7
LIST OF TABLES .....	8
Acronyms.....	9
1. Executive summary.....	10
Context and purpose of the meta-evaluation .....	16
2.1. Context.....	16
2.2. Meta-evaluation background, scope and purpose .....	16
2.3. Approach .....	18
2.4. Methodology.....	19
a. Data sources .....	24
b. Data analysis approach .....	24
c. Sampling procedures .....	25
d. Cross-cutting issues.....	27
e. Norms and standards .....	28
2.5. Limitations of the meta-evaluation .....	28
Findings on evaluation report quality .....	29
3.1. Overall findings on evaluation report quality .....	31
3.2. Geographical analysis.....	33
3.3. Analysis by evaluator .....	35
3.4. Analysis by thematic area .....	38
3.5. Analysis per evaluation scope.....	40
3.6. Analysis per evaluation budget.....	43
3.7. Analysis by component .....	44
1 Component 1. Executive summary .....	45
2 Component 2. Project background .....	46
3 Component 3. Evaluation background .....	47
4 Component 4. Evaluation criteria and questions .....	48
5 Component 5. Evaluation methodology .....	49
6 Component 6. Evaluation findings .....	51
7 Component 7. Conclusions .....	52
8 Component 8. Recommendations .....	53
9 Component 9. Lessons learned .....	54
10 Component 10. Good practices .....	55
3.8. Comprehensiveness of the evaluation reports.....	56
3.9. Comprehensiveness of the terms of reference .....	58
3.10. Comprehensiveness of inception report .....	62
Review by sections .....	62
Overall analysis.....	64
Findings on the evaluation process.....	64
4.1 Evaluation design .....	69
1.2 Evaluation implementation.....	71
Findings on evaluation capacities and skills.....	73
5.1. Internal evaluator training .....	73
5.2. Use of key evaluation documents.....	74

5.3. Self-perceived preparedness of internal evaluators.....	76
Conclusions .....	77
Good practices .....	81
Lessons learned .....	82
Recommendations .....	82
ANNEX I: Analytical framework - Dimension 1 - Quality of evaluations.....	88
ANNEX II: Analytical framework - Dimension 2 - Comprehensiveness of evaluation reports.....	92
ANNEX III: Analytical framework - Dimension 3 - Comprehensiveness of evaluation terms of reference	
93	
ANNEX IV: Strategies to ensure consistency.....	94
ANNEX V: List of documents reviewed .....	99
ANNEX VI: List of evaluations.....	101
ANNEX VII: List of terms of reference and inception report sampling .....	113
ANNEX VII: Global scores of the components by geographic area.....	115
ANNEX IX: List of people interviewed .....	117
ANNEX X: Interview protocol .....	117
ANNEX XI: Survey protocol - Evaluation managers.....	119
ANNEX XII: Survey protocol evaluators (internal and external) .....	123
ANNEX XIII: Terms of reference .....	128
ANNEX XIV: Inception report .....	133

## LIST OF FIGURES

FIGURE 1. RELATIVE WEIGHT FOR THE RATING OF EACH COMPONENT .....	24
FIGURE 2. OVERVIEW OF EVALUATIONS UNDER ASSESSMENT .....	26
FIGURE 3. OVERALL PROCESS OF THE META-EVALUATION .....	27
FIGURE 4. EVOLUTION OF GLOBAL RATING 2017–2024 PER LEVELS .....	32
FIGURE 5. DISTRIBUTION OF REPORTS PER RATING 2020–2024 .....	32
FIGURE 6. EVOLUTION OF MEAN GLOBAL RATING 2017–2024 .....	32
FIGURE 7. GLOBAL RATING LEVELS PER REGION.....	34
FIGURE 8. GLOBAL MEAN RATINGS PER REGION (2017–2019 AND 2020–2024) .....	35
FIGURE 9. BOX DIAGRAM OF EVALUATION RATING BASED ON THE TYPE OF EVALUATOR .....	36
FIGURE 10. EVOLUTION OF AVERAGE GLOBAL RATING PER COMPONENT AND TYPE OF EVALUATOR.....	37
FIGURE 11. GLOBAL AVERAGE RATING PER THEMATIC AREA.....	39
FIGURE 12. COMPARISON OF GLOBAL RATINGS PER THEMATIC AREA AND PERIODS.....	40
FIGURE 13. COMPARISON OF GLOBAL RATINGS PER GEOGRAPHICAL COVERAGE .....	41
FIGURE 14. COMPARISON OF GLOBAL RATINGS BY MODALITY .....	42
FIGURE 15. COMPARISON OF GLOBAL RATINGS PER TIMING.....	42
FIGURE 16. CORRELATION MODEL FOR EVALUATION RATING BASED ON PERCENTAGE OF THE OVERALL BUDGET ALLOCATED TO EVALUATION.....	43
FIGURE 17. ARITHMETIC MEANS OVERALL RATING PER COMPONENT .....	44
FIGURE 18. COMPREHENSIVENESS OF EVALUATION REPORTS (N = 90) .....	57
FIGURE 19. COMPREHENSIVENESS OF TERMS OF REFERENCE (N = 58).....	59
FIGURE 20. ASSESSMENT OF THE DIFFERENT SECTIONS OF THE TERMS OF REFERENCE.....	60
FIGURE 21. RELATIONSHIP BETWEEN GLOBAL SCORING TERMS OF REFERENCE AND GLOBAL SCORING REPORTS.....	61
FIGURE 22. GLOBAL AVERAGE QUALITY SCORE OF THE TERMS OF REFERENCE BY YEARS .....	61
FIGURE 23. LEVEL OF AGREEMENT OF SURVEYED EVALUATION MANAGERS WITH STATEMENTS ABOUT THE EVALUATION PROCESS .....	66
FIGURE 24. KEY CHALLENGES PERCEIVED BY SURVEYED EVALUATORS.....	68
FIGURE 25. NUMBER AND PERCENTAGE OF PARTICIPANTS IN MONITORING AND EVALUATION TRAINING .....	74
FIGURE 26. PROPORTION OF STAKEHOLDERS FREQUENTLY OR ALWAYS USING KEY EVALUATION DOCUMENTS .....	76
FIGURE 27. PROPORTION OF STAKEHOLDERS CONSIDERING THEMSELVES SOMEWHAT OR HIGHLY PREPARED TO CONDUCT EVALUATIONS .....	77

## LIST OF TABLES

TABLE 1. INTENDED USERS AND EXPECTED USE FOR THE META-EVALUATION .....	18
TABLE 2. MAIN AREAS OF ANALYSIS OF THE META-EVALUATION .....	18
TABLE 3. METHODOLOGICAL APPROACH.....	19
TABLE 4. META-EVALUATION QUESTIONS AND TECHNIQUES .....	20
TABLE 5. DIMENSIONS UNDER ANALYSIS THROUGH THE META-EVALUATION TOOL.....	21
TABLE 6. OVERALL SCORING SYSTEMS .....	22
TABLE 7. DATA SOURCES: EXISTING DOCUMENTS .....	24
TABLE 8. RISKS, LIMITATIONS AND MITIGATION STRATEGIES.....	28
TABLE 9. EVOLUTION OF GLOBAL RATING PER THEMATIC AREA .....	39
TABLE 10. OVERALL RESULTS FOR COMPONENT 1, EXECUTIVE SUMMARY.....	45
TABLE 11. OVERALL RESULTS FOR COMPONENT 2, PROJECT BACKGROUND.....	46
TABLE 12. OVERALL RESULTS FOR COMPONENT 3, EVALUATION BACKGROUND .....	47
TABLE 13. OVERALL RESULTS FOR COMPONENT 4, EVALUATION CRITERIA AND QUESTIONS .....	48
TABLE 14. OVERALL RESULTS FOR COMPONENT 5, EVALUATION METHODOLOGY.....	49
TABLE 15. OVERALL RESULTS FOR COMPONENT 6, EVALUATION FINDINGS .....	51
TABLE 16. OVERALL RESULTS FOR COMPONENT 7, CONCLUSIONS.....	53
TABLE 17. OVERALL RESULTS FOR COMPONENT 8, RECOMMENDATIONS.....	54
TABLE 18. OVERALL RESULTS FOR COMPONENT 9, LESSONS LEARNED.....	55
TABLE 19. OVERALL RESULTS FOR COMPONENT 10, GOOD PRACTICES.....	56

## Acronyms

<b>ALNAP</b>	Active Learning Network for Accountability and Performance
<b>CED</b>	Central Evaluation Division
<b>DAC</b>	Development Assistance Committee
<b>IOM</b>	International Organization for Migration
<b>M&amp;E</b>	Monitoring and evaluation
<b>MOPAN</b>	Multilateral Organisation Performance Assessment Network
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>OIG</b>	Office of the Inspector General (renamed as OIO)
<b>OIO</b>	Office of Internal Oversight
<b>PRIMA</b>	PRIMA
<b>RBA</b>	Human rights-based approach
<b>ROMEO</b>	Regional monitoring and evaluation officers (now RPMEO)
<b>RPMEO</b>	Regional planning, monitoring and evaluation officers
<b>ToR</b>	Terms of reference
<b>UN-SWAP</b>	United Nations System-wide Action Plan
<b>UNEG</b>	United Nations Evaluation Group

## 1. Executive summary

This report presents the findings of a meta-evaluation of evaluations commissioned by the International Organization for Migration (IOM) between 2020 and 2024 (hereafter referred to as “the meta-evaluation”). The meta-evaluation was carried out by Artival Research & Evaluation between November 2024 and April 2025. Its main objectives were:

1. To assess the quality of internal and external evaluations (centralized and decentralized)<sup>1</sup> conducted between 2020 and 2024 and provide actionable recommendations to enhance the quality and utilization of evaluations.
2. Assess the relevance and use of the quality assessment tools introduced, analysing opportunities and limitations on their use and identifying other relevant factors that influence the use of evaluation results.
3. To assess the evolution of the evaluation quality compared with the 2017–2019 meta-evaluation.

The primary users of this meta-evaluation are IOM evaluation staff, including regional and country officers responsible for evaluation, who are tasked with implementing quality management mechanisms and providing support and guidance. This evaluation also benefits IOM senior management by providing recommendations on how to improve the quality of evaluation products, ultimately improving the utilization of evaluation recommendations to enhance programming and decision-making.

### Methodology

The meta-evaluation adopted a utilization-focused approach with a mixed-method concurrent triangulation strategy, combining quantitative and qualitative data to enhance the validity of findings. The evaluation questions focused on assessing the quality of IOM evaluations, the evaluation processes and capacities, and the use of evaluation tools. The complete list of questions is presented in Table 4.

The assessment examined a sample of 90 evaluation reports with particular attention to comprehensiveness, clarity and alignment with institutional standards. Additionally, it assessed the quality a subset of the associated ToRs<sup>2</sup> and a limited number of inception reports.<sup>3</sup>

The methods employed included an evaluation scoring tool, considering factors like budgets, internal versus external conduct of evaluations, geographical scope and type of evaluation. Statistical methods, including correlation coefficients, were applied to identify patterns and visualize trends. Semi-structured interviews with 14 internal and external stakeholders, and three online surveys were conducted to further support the triangulation of insights, one targeting internal evaluators, one targeting external evaluators and one targeting evaluation managers. The assessment considered whether and how the design of evaluations addressed gender equality, RBA in programming, disability inclusion and environmental sustainability, as

---

<sup>1</sup> Centralized evaluations are commissioned by CED. Decentralized evaluations include those commissioned by any other IOM entity, irrespective of the level (global, regional, or national).

<sup>2</sup> The comprehensiveness of the ToRs was assessed for all 58 ToRs attached to the sampled evaluation reports, while the quality assessment was conducted on a subset of 19 reports.

<sup>3</sup> Three inception reports selected from a pool of 25 inception reports available in the sampled evaluation reports.

well as how and whether these issues were addressed in the findings, conclusions and recommendations of evaluation reports.

Limitations encountered included the absence of systematically recorded data on evaluator training and evaluation budgets, which constrained the ability to robustly analyse the relationship between training, budget allocation and evaluation quality. Additionally, the inclusion of 2024 data represented a limitation due to the small sample size, necessitating cautious interpretation of the findings as preliminary trends rather than definitive conclusions.

### **Key findings**

Concerning the quality of evaluations, the 2020–2024 assessment shows a moderate improvement, with 39 per cent of reports meeting or exceeding requirements, 44 per cent approaching requirements, and 17 per cent falling below standards. The global average score rose from 2.6 in the 2017–2019 cycle to 2.8 in the 2020–2024 cycle (with a maximum score of 5).

The proportion of budget allocated to evaluation, rather than its absolute size, was significantly linked to higher quality outcomes. Components like “findings” and “evaluation criteria” outperformed others, while “lessons learned” and “good practices” remained weak points.

The quality of terms of reference (ToRs) positively influenced overall evaluation quality, highlighting the need for strong methodology and comprehensive cross-cutting analysis. However, ToRs presented inconsistencies in terms of clarity and relevance, frequently applying all OECD-DAC criteria without assessing contextual applicability. Despite improvements, gaps remain in addressing disability inclusion and environmental sustainability, and in documenting emerging good practices.

When it comes to the evaluation process, the decentralized approach to evaluation design within IOM has led to variability in scope, methodological rigour and consistency. While institutional policies and donor requirements offer general guidance, evaluations are often planned ad hoc, especially for smaller or short-term projects, with limited systematic oversight. The absence of a standardized review process for decentralized evaluations and the irregular involvement of RPMEOs further contributed to uneven quality.

Although quality assurance tools – like templates and checklists – have introduced greater structure, their application remains inconsistent across offices, often hindered by competing priorities or limited awareness. Furthermore, inconsistencies in data collection and informant selection compromise analytical depth and increase the risk of bias.

Structural weaknesses persist in institutionalizing the evaluation function, characterized by fragmented responsibilities, and a lack of mandatory procedures for reviewing evaluation deliverables for decentralized evaluations. Recent initiatives, such as wider dissemination of evaluation briefs and donor engagement, offer promising avenues for strengthening learning, strategic alignment and the use of evaluation findings in decision-making processes.

Finally, regarding evaluation capacities and skills, internal evaluation training within IOM sharply declined after peaking in 2021 with 256 participants, dropping to just 1 by 2024. While

core resources like the *IOM Project Handbook* (2017) and the *IOM Monitoring and Evaluation Guidelines* (2021) remain widely used, inconsistent consultation of critical documents – such as the “Guidance for addressing gender in evaluations” (2018) – may contribute to the gaps in addressing cross-cutting issues.

## Conclusions

1. The evaluation quality trajectory – marked by a peak in 2020, followed by a decline to its lowest point in 2022, and a subsequent moderate recovery in 2023 – indicates resilience in the evaluation function beyond pandemic effects. This disruption of project objectives, limited access to key informants, and complicated evaluation coordination has left a lasting effect. Reports from 2020–2024 indicate extended deadlines and modified methodologies. While the evaluation guidelines set a framework for improvement, their impact on report quality remains unclear due to limited use. While there are, since 2022, signs of improvement in the quality of evaluations, methodological gaps, limited adaptive strategies and inconsistent quality assurance processes persist.
2. While evaluations generally “approach requirements” and certain evaluation components are generally of high quality, there are critical gaps, particularly in integrating cross-cutting perspectives, refining methodologies and knowledge management, especially regarding lessons learned and good practices. Across evaluations, the quality of project backgrounds, recommendations and executive summaries generally meet the minimum quality standards, though there is room for improvement and need for attention.
3. Regional offices have played a crucial role in enhancing evaluation quality by providing practical technical guidance and on-the-spot capacity-building initiatives, particularly through comprehensive methodological support. The early involvement of regional offices in drafting ToRs and strategic midterm evaluations have facilitated real-time learning and adjustments. Furthermore, evaluation quality is highest in regions where RPMEOs are actively involved in the review of ToRs and key deliverables, suggesting that institutional support is a key determinant of methodological rigour. Evaluation reports often demonstrate deeper analysis and stronger adherence to institutional standards when internal evaluators and managers are consistently guided.
4. Inadequate time allocation for evaluation processes has led to tight timelines and superficial analysis, limiting the depth and rigour of assessments. Delays in commissioning evaluations and short time frames hindered data collection and analysis, often leading to last-minute adjustments and methodological compromises. As a result, many reports prioritize compliance with donor requests over meaningful learning. Although the Organization’s evaluation culture has improved since 2020, evaluations are still largely seen as contractual obligations rather than an integral tool for learning and decision-making.
5. The analysis shows that the proportion of the total budget dedicated to evaluation significantly influences evaluation quality. This underscores the need to dedicate an appropriate percentage of the project budget to evaluation rather than focusing on absolute

amounts. In the current context of budget constraints,<sup>4</sup> it is crucial to ensure, in the initial formulation, that a dedicated percentage of the project budget is earmarked for evaluation.

6. Despite the challenges in ensuring methodological rigour and consistent quality assurance, several practices have enhanced evaluation processes within IOM. Tools like PRIMA and proactive measures by regional offices, such as quarterly trackers and direct follow-ups, have increased transparency and improved preparation, resulting in more structured evaluations. Comprehensive training materials have strengthened evaluation capacities and culture. The use of standardized templates and quality assurance tools has reduced inconsistencies, thus increasing the credibility of findings. Some regional offices and RPMEOs have developed effective practices with the potential to support broader institutional learning. Effective evaluation practices, including targeted efforts to address existing gaps, have the potential to elevate the quality and impact of IOM evaluations, ultimately supporting strategic decision-making and organizational advancement.

7. The introduction of standardized indicators (through the Strategic Results Framework) and the push to utilize monitoring mechanisms have helped to reinforce the connection between evaluations and institutional priorities. This stronger alignment has occasionally led to more organized assessments of programme outcomes, thereby boosting the ability of evaluations to guide future programming and policy decisions.

8. Decentralization without standardized and clear procedures for design, approval and quality control has led to inconsistent evaluation quality, often dependent on the experience and engagement of individual PMs. The absence of systematic oversight, reliance on voluntary compliance, and unclear roles and responsibilities further weaken quality assurance. RPMEOs are underutilized in evaluation processes and disproportionately tasked with strategic planning and monitoring; the challenges they face are compounded by structural limitations, scarce resources and tight timelines. Institutional mechanisms for empowering M&E officers remain vague, hindering consistent evaluation practices and alignment with IOM strategic goals. Strengthening these mechanisms would significantly improve evaluation quality and utility while fostering a culture of learning, adaptation and continuous improvement across the Organization.

9. Inconsistent promotion of quality assurance mechanisms and a lack of methodological rigour undermine the credibility and depth of evaluations. Weak sampling strategies, insufficient triangulation and inadequate traceability and transparency in data collection and analysis often lead to superficial findings and fail to provide robust evidence for conclusions and recommendations. Many reports fail to establish a clear and logical link between analysis and proposed actions, leading to recommendations that are too broad and not evidence based. The absence of structured stakeholder engagement and participatory discussion in the recommendation process further weakens their relevance and applicability.

10. Well-crafted ToRs are positively correlated with more robust evaluation reports. Although the “Guidance on quality management of IOM evaluations” (2022) laid a foundation for enhancing the quality of ToRs, many ToRs continue to exhibit inconsistencies and gaps. The

---

<sup>4</sup> For instance, the budget adjustments resulting from the United States realignment of foreign aid. See the [presidential action reevaluating and realigning United States foreign aid](#) (20 January 2025).

blanket inclusion of all OECD-DAC criteria and the lack of contextual adaptation leads to broad scopes and diluted analysis. Structural factors, such as the expertise of the evaluation team and managerial support, also influence outcomes. Regional offices, often reliant on PMs' individual initiative, coupled with high staff turnover and limited technical capacity in some country offices, further undermine consistency. Addressing these gaps is essential to sustaining and enhancing the quality of ToR and, consequently, the overall effectiveness of evaluations.

11. The assessment of the usefulness of IOM methodological guidelines and quality assurance tools reveals a varied and complex reality across different stakeholders. This suggests that their effectiveness and application differ based on context, user familiarity and the specific needs of the evaluation, rather than providing a uniform or consistently applied approach. These guidelines and tools offer a robust framework that evaluation managers appreciate for its structured approach. However, their complexity, length and inconsistent adoption pose significant challenges in practical application. The discontinuity of training, frequent staff turnover and lack of structured follow-up mechanisms hinder their consistent and effective use.

12. The inconsistent uptake of quality assurance tools signals that voluntary compliance is insufficient. Embedding quality assurance checks into mandatory approval workflow would promote uniformity, reliability and methodological rigour. Requiring the use of quality assurance tools as a condition for evaluation approval helps guarantee that all evaluations meet established quality benchmarks before moving forward. This proactive approach not only enhances the quality and credibility of evaluation outputs but also fosters a culture of accountability and continuous improvement. In the absence of such structured enforcement, quality assurance tool usage is likely to remain irregular, increasing the risk of inconsistency, inefficiency and reduced organizational effectiveness.

13. Persistent gaps in technical capacity – exacerbated by high turnover and sporadic training – highlight a systemic challenge in institutionalizing evaluation skills and in applying best practices. Building a resilient evaluation function will require long-term investment in structured learning pathways, rather than ad hoc efforts.

### **Good practices**

The active involvement of RPMEOs has played a crucial role in enhancing the quality of evaluations, especially in terms of methodological rigour and coherence. Their oversight, particularly in reviewing ToRs, methodologies and final reports, has led to stronger evaluation reports with clearer recommendations. However, the effectiveness of this role varies widely across regions, as some country offices do not systematically involve RPMEOs in the evaluation process.

Internal evaluation databases facilitate knowledge sharing and inform new projects, but accessibility challenges limit their effectiveness, as not all staff are aware of these resources or how to utilize them effectively.

Adopting standardized evaluation tools, particularly through structured templates and checklists, has enhanced consistency in reporting, streamlining the documentation process and ensuring the inclusion of key components such as methodology, findings and recommendations. While this standardization does not guarantee content quality, it has significantly improved the coherence and comparability of evaluation reports across different offices.

### **Lessons learned**

Integration between monitoring and evaluation functions remains weak, with many assessments relying on weak or incomplete monitoring data, often due to poorly designed results matrices and limited indicators that fail to capture key outcomes.

Technical evaluation expertise gaps among some evaluation managers and programme staff, particularly in country offices lacking dedicated M&E staff, have undermined methodological rigour, leading to inconsistencies in data collection, weak sampling strategies and a lack of analytical depth in findings.

Inconsistent institutional and donor requirements for evaluations have led to their treatment as optional formalities rather than essential project management and learning tools.

### **Recommendations**

The recommendations summarized here are further detailed in the section on recommendations, below. Each is an action that IOM should undertake.

1. Keep promoting the use of the evaluation report templates and ensure that their use is mandatory to ensure that all the required sections of such reports are included.
2. Continue advocating for the inclusion of gender equality and RBA in programming as cross-cutting issues as well as promoting environmental sustainability and disability inclusion during evaluations. This integration should begin with early needs assessments, stakeholder identification, and the development of the evaluation methodology, continuing through the findings and recommendations sections.
3. Continue promoting that a fixed percentage of the overall project budget be allocated for evaluation to ensure the quality of evaluations, as recommended in the *IOM Monitoring and Evaluation Guidelines*. This approach appears to ensure evaluations have the necessary resources to be of high quality.
4. With regard to the IOM evaluation framework, promote a culture of adaptation that enables methodological adjustments tailored to the specific context of each evaluation, in order to improve the utility and impact of the framework.
5. Enhance the quality of ToRs by strengthening the clarity, focus and methodological rigour, and by establishing review procedures, ultimately leading to more impactful and actionable insights.

6. Make the quality assurance tools more accessible and practical to promote their consistent use across different teams.

7. Build long-term capacity in evaluation through structured and continuous training efforts. Relying on one-off webinars or ad hoc learning opportunities has proven insufficient, particularly considering high staff turnover.

## Context and purpose of the meta-evaluation

### 2.1. Context

IOM is the leading intergovernmental organization in the field of migration within the United Nations system, dedicated to promoting humane and orderly migration since 1951. Present in 171 countries and operating over 550 field offices worldwide across six regions, IOM implements over two thousand projects per year, across different thematic areas.

Evaluation is an essential function for the effective achievement of the IOM mandate and the efficient implementation of its strategic plan. It ensures accountability and evidence-based oversight of performance and results, informs decision makers and supports institutional learning.

This second meta-evaluation reviewed IOM evaluations from 2020 to 2024. The main beneficiaries of this evaluation include the CED and RPMEOs tasked with implementing quality management mechanisms and providing support and guidance, as well as internal and external evaluators and PMs. The meta-evaluation supports IOM senior management by, in the long term, enhancing the quality of evaluation reports, thereby enabling the Organization to utilize evaluation findings, conclusions and recommendations to enhance programming and decision-making. The meta-evaluation also aims to provide insights to secondary users – including other United Nations agencies and organizations, civil society, Member States, donors and assisted populations – on the quality of IOM evaluations, thus fostering transparency and ensuring accountability.

### 2.2. Meta-evaluation background, scope and purpose

The ToR of the meta-evaluation emphasize that evaluation is important for learning, accountability, adaptation and evidence-based decision-making (Annex XIII). The IOM evaluation function works on two levels: central evaluations, which encompass corporate and strategic evaluations, are overseen and managed by the CED; and decentralized evaluations, which primarily focus on programme and project evaluations, are typically commissioned by team leaders at the country, regional, or Headquarters levels.

After the [MOPAN 2017–18 assessment of IOM](#), the Organization conducted the first meta-evaluation of its internal and external evaluations from 2017 to 2019. The resulting meta-evaluation, presented in 2020, stated that the regular conduct of such meta-evaluation exercises in IOM deserves to be formally established to allow an analysis of the quality of evaluation.

This exercise, together with the “[UNEG/OECD peer review of the International Organization for Migration \(IOM\) evaluation function](#)” (2021), concluded with recommendations to improve the quality of IOM evaluations. The recommendations included:

1. Introducing a quality assurance system for central and decentralized evaluations, ensuring that there is quality assurance for terms of references, inception, draft and final evaluation reports.
2. Introducing an external post-hoc quality assessment of all evaluations.
3. Encouraging the use of evaluation reference groups.

IOM has fostered a culture of evaluative understanding and practice to enhance the quality assurance system. This is promoted by acknowledging the importance of evaluations in reporting intervention outcomes and lessons learned, as well as identifying innovative and replicable practices.

Since 2020, IOM has been working to enhance its organizational evaluation function through a feasibility study on a quality management mechanism (2022), including a review of systems employed by other United Nations agencies and a new selection of evaluation reports released in 2020 and 2021. To ensure adherence to all quality requirements, several guidelines have been approved: “[Guidance for evaluation managers](#)” (2021), “[Guidance for evaluators](#)” (2021), “[Guidance on quality management of IOM evaluations](#)” (2022), and “[Guidance on the use of evaluations and follow-up of evaluation recommendations](#)” (2022), along with IOM evaluation strategies (2021–2023 and [2024–2028](#)). A review of the use and follow-up of management responses in IOM has been carried out and was completed by February 2025.<sup>5</sup> In this context, this second meta-evaluation, foreseen in the CED evaluation strategy 2024–2028, assessed the quality of IOM evaluation reports with three main objectives:

1. To assess the quality of internal and external evaluations (centralized and decentralized) conducted between 2020 and 2024 and provide actionable recommendations to enhance the quality and utilization of evaluations.
2. Assess the relevance and use of the quality assessment tools introduced, analysing opportunities and limitations on their use and identifying other relevant factors that influence the use of evaluation results.
3. To assess the evolution of the quality of evaluations, compared with the quality of those assessed in the 2017–2019 meta-evaluation.

The meta-evaluation assessed internal and external evaluation reports published between 2020 and 2024 (90 reports). Additionally, it assessed a subset of the associated ToRs (19 ToRs) and a limited number of inception reports (3 reports). The purposive sample was derived from a pool of 265 evaluations using random sampling, then systematically adjusted to ensure balanced representation across different variables including year of publication, type of evaluator and evaluation coverage. (More information about the sampling strategy and the evaluation reports assessed can be found in section 2.4.c, and especially in Figure 2.)

The meta-evaluation focused on maximizing the utility of its results and their actual use. To this end, the design and process carefully considered meeting the information needs of primary and secondary audiences, as tentatively expressed in Table 1.

---

<sup>5</sup> See Arild Hauge, “[Review of the use and follow-up of \(evaluation\) management response in IOM](#)” (Oslo, 4 February 2025).

Table 1. Intended users and expected use for the meta-evaluation

Type of user	Designation	Intended use
Primary	Central Evaluation Division Regional or country level planning and M&E officers Evaluation managers Internal evaluators	<ul style="list-style-type: none"> <li>▪ Strengthening of IOM evaluation function</li> <li>▪ Improving preparation, implementation and resource management for evaluation</li> <li>▪ Accountability</li> <li>▪ Learning</li> </ul>
	IOM senior management and IOM in general	<ul style="list-style-type: none"> <li>▪ Improving the quality and use of evaluation reports</li> <li>▪ Improving practices in the future as IOM staff manage, conduct, or use evaluations</li> </ul>
Secondary	Other United Nations agencies and organizations	<ul style="list-style-type: none"> <li>▪ Organizational learning and development</li> </ul>
	Indirect beneficiaries and civil society	<ul style="list-style-type: none"> <li>▪ Learning and development, accountability</li> </ul>
	Member States and donors	<ul style="list-style-type: none"> <li>▪ Accountability</li> </ul>

The primary users of this meta-evaluation are the IOM evaluation staff, including the regional and country officers responsible for evaluation, who are tasked with implementing quality management mechanisms and providing support and guidance. Additionally, this meta-evaluation benefits IOM senior management by providing recommendations on how to improve and make better use of evaluations, ultimately improving the utilization of evaluation recommendations to enhance programming and decision-making. The meta-evaluation is also important for secondary users (other United Nations agencies and organizations, civil society, Member States, donors and beneficiaries of IOM interventions, providing insights into the quality of IOM evaluations, thus fostering transparency and ensuring accountability.

### 2.3. Approach

The meta-evaluation used both quantitative and qualitative approaches concurrently to triangulate findings, comparing the data sets produced by each approach to identify convergences, divergences, or a combination of both. This process included confirmation, disconfirmation, cross-validation and corroboration to derive key findings that informed the conclusions and recommendations. In practice, quantitative data emphasized statistical trends, while the qualitative data gathered during the assessment provided explanations and nuances for these trends.

The terms of reference for this meta-evaluation listed nine questions, which can be summarized into three driving questions: “What is the quality of the evaluations conducted? How has evaluation quality changed, especially as compared to 2017–2019? And what is needed to strengthen the IOM evaluation function in the future?” These questions were then used to identify the main areas of analysis, as summarized in Table 2.

Table 2. Main areas of analysis of the meta-evaluation

Key concern	Areas of analysis
Evaluation quality	<ul style="list-style-type: none"> <li>▪ Overall evaluation quality and ways to improve it</li> <li>▪ Strengths and weaknesses of evaluation processes</li> <li>▪ Influence of cost and timing on the overall quality of evaluation</li> <li>▪ Evolution in the quality of evaluation since 2017</li> </ul>

Key concern	Areas of analysis
Evaluation capacities	<ul style="list-style-type: none"> <li>▪ Evaluation quality per type of evaluator (internal and external)</li> </ul>
Use of evaluation tools	<ul style="list-style-type: none"> <li>▪ Use of IOM quality control tools for ToRs and inception reports</li> <li>▪ Issues that facilitated or hindered the use of evaluation tools</li> <li>▪ Factors that influenced the use of evaluation results</li> </ul>

## 2.4. Methodology

The methodological strategies and data-gathering techniques considered are laid out in Table 3.

Table 3. Methodological approach

Methodological strategies	Research techniques	Data gathering tools
<ul style="list-style-type: none"> <li>▪ Qualitative strategies (discourse analysis)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Semi-structured interviews</li> <li>▪ Document review</li> </ul>	<ul style="list-style-type: none"> <li>▪ Interview protocols</li> <li>▪ Scoring tool</li> </ul>
<ul style="list-style-type: none"> <li>▪ Quantitative strategies (descriptive and inferential statistical analysis)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Analytical framework (desk review)</li> <li>▪ Online survey</li> </ul>	<ul style="list-style-type: none"> <li>▪ Questionnaires</li> <li>▪ Scoring tool</li> </ul>

The methodological proposal intended to solve methodological challenges by:

- Combining quantitative and qualitative data analysis. Quantitative data gathered from primary and secondary sources (such as the scoring tool, the online survey and document reviews) provided substantial information on effectiveness, while data gathered through qualitative methods (interviews, document reviews and the scoring tool's open-ended field) provided meaningful information about all components and notably on the utility and relevance of the evaluations and ToR produced.
- Using different and complementary sources. The proposed approach used different sources to fill gaps that arose due to information being insufficient, outdated, or unreliable. Data obtained from the assessment of reports, ToRs and inception reports was complemented by individual opinions (structured interviews and surveys). Different sources allowed for the triangulation of information, especially when searching for good practices and recommendations for the future.

The four complementary research techniques considered were:

1. Assessment with the scoring tool.<sup>6</sup> The analytical framework and its scoring tool provided essential data with which to assess the quality of evaluation reports, allowing the emergence of statistical trends by utilizing statistical association through correlation coefficients. The analysis examined whether evaluation budgets and the conduct of evaluations (internal or external) correlated with quality ratings.<sup>7</sup> Additionally, it considered factors like geographical area, type of evaluation and coverage. Once the quantitative database was prepared, specific software, including Python and Excel, was used for analysis and to create visuals.
2. Assessment of the quality of the ToRs and inception reports. The ToR analysis employed a dual approach to thoroughly evaluate their content and structure. A dichotomous assessment, using the meta-evaluation scoring tool, systematically checked for the presence of key components in the ToR with a "yes/no" scale. This tool assessed

<sup>6</sup> The scoring tool from the initial meta-evaluation has been revised to better capture the specific goals of this assessment. The scoring tool used to assess each of the three dimensions can be found in annexes I–III, and the strategies used to ensure consistency can be found in Annex IV.

<sup>7</sup> When the evaluation budget was not available, the evaluation expenditures were used instead.

whether the ToR contained essential elements such as background context, evaluation purpose, scope and goals, criteria and questions, methodology, cross-cutting issues, deliverables and budget, among others.

In parallel, an analysis was conducted following the IOM guidelines.<sup>8</sup> For this analysis, each ToR was evaluated using a structured scoring system based on the 13 sections or components outlined in the IOM checklist.<sup>9</sup> Each section was assessed based on the elements that defined its expected content, and all sections were given equal weight to ensure consistent scoring. Equal weighting was selected because each component equally contributes to the overall quality and usability of the ToR.

The final score for each ToR was determined based on the number of required elements (comprehensiveness of ToR section) and their quality in terms of clarity, specificity and alignment with IOM guidelines. This approach allowed for a detailed, structured evaluation of ToRs, ensuring that both the presence and the adequacy of each section were considered.

In addition, a small sample of inception reports were analysed, using the IOM “[Quality control tool – inception reports](#)” (2022), as the guideline for assessing their quality.

3. Semi-structured interviews. Interviews with a select number of IOM staff and external evaluators were a crucial method of information-gathering in the meta-evaluation, allowing the triangulation of information, points of view and opinions of different stakeholders. The views and information deepened the findings of the analysis. A total of 14 interviews were conducted, with country and regional planning, monitoring and evaluation officers and with a selection of key informants suggested by CED and the reference group. The list of people interviewed is available in Annex IX.
4. Survey. An online survey was developed and launched to complement qualitative information with related quantitative data. The questionnaires were sent to internal and external evaluators and the managers of the evaluations conducting or managing evaluations during the period.

The consulting team developed the tools concurrently (see annexes X, XI and XII). These were informed by the nine evaluation questions derived from those given in the meta-evaluation ToR, and the key areas under assessment included in the scoring tool.

Table 4. Meta-evaluation questions and techniques

Meta-evaluation questions <sup>10</sup>	Concept	Research tools
1. To what extent do IOM evaluation reports and their ToRs meet the IOM and UNEG evaluation quality standards as per the IOM quality assessment criteria?	Quality of evaluations	Scoring tool Interview
2. On which quality-related aspects or factors do the evaluation processes (from preparation to reporting) and evaluation reports excel and on which are they falling short?	Quality of evaluations	Scoring tool Interview Survey
3. To what extent is there evidence on the use of the quality control tools outlined in the “ <a href="#">Guidance on quality management of IOM evaluations</a> ” (2022)? <sup>11</sup>	Internal use of evaluation tools	Scoring tool Interview Survey

<sup>8</sup> IOM Central Evaluation Unit, “[Guidance on quality management of IOM evaluations](#)” (September 2022).

<sup>9</sup> Ibid.

<sup>10</sup> Some of the original meta-evaluation questions from the ToR were consolidated to make them more straightforward and streamlined, ensuring that all the main aspects were still included.

<sup>11</sup> The original question 3.3, “What other factors influence the use of evaluation results?” has been excluded, as it remains outside the scope of this assessment.

Meta-evaluation questions <sup>10</sup>	Concept	Research tools
3.1. What factors facilitated or hindered the use of the tools to ensure the quality of evaluations? 3.2 How adequate are the quality assurance tools developed by CED?		
4. What changes can be observed in terms of the quality of evaluations as compared with the quality of the evaluations assessed in the previous meta-evaluation, and what factors (beyond the use of quality assurance tools) may have contributed towards these changes?	Evolution of the quality of evaluation reports	Scoring tool Interview
5. Have the costs and resources assigned to the evaluations positively correlated with the quality of the final product?	Quality of evaluations	Scoring tool Interview Survey
6. What are the main differences between evaluations conducted by internal evaluators and those conducted by external evaluators?	Quality vs. type of evaluators	Scoring tool Interview Survey

Table 4 shows that the scoring tool played a major role in the analysis. Developed for the previous meta-evaluation, it was revised to address the specificities of this assignment. It was based on the meta-evaluation questions, and a series of rubrics and ratings were specifically developed to assess the extent to which each concept or evaluation question was addressed.

The analytical framework that served as the key tool to conduct the meta-evaluation used four key terms to frame its analysis: dimensions, components, items and elements.

- Dimensions are at the broadest level the things being assessed, and each is assessed differently. This meta-evaluation assessed three dimensions: (1) the quality of the evaluation reports; (2) the comprehensiveness of the evaluation reports, and (3) the comprehensiveness of the ToRs of the evaluation reports (see Table 5).
- Components are the aspects that each dimension is expected to possess or demonstrate. For the most part, they are equivalent to the sections of the report or of the ToR. For example, in assessing the comprehensiveness of the ToRs (dimension 3), it was expected that each ToR would have certain key sections such as evaluation background, methodology, findings and recommendations. .
- Items are the key requirements that each evaluation report, inception report or ToR is expected to manifest. For example, an evaluation report's project background is expected to offer a brief outline of contextual factors, a clear description of key stakeholders, a description of the project logic and a summary of funding arrangements. The quality of the component is determined by whether and how well it presents those items.
- Elements are content specifications that define what must be present to satisfy the completeness of each assessed item. For instance, to continue the example above, the assessment of the description of key stakeholders (item) of the project background (component) of each evaluation report would be made based on whether and how well each of the relevant rights holders, implementing agencies, and duty bearers or responsibility holders were identified.

Table 5. Dimensions under analysis through the meta-evaluation tool

Dimension	Analysis mechanism
-----------	--------------------

1. Quality of evaluation reports	Collect information about the quality of evaluation reports. Rate the quality and relevance of the information within reports.
2. Comprehensiveness of evaluation reports	Produce a summary and checklist of the information contained in the evaluation reports.
3. Comprehensiveness of ToR	Assess the key sections and information contained in the terms of reference and inception reports.

Dimension 1 is the most comprehensive and provides useful information for analysing ten components: the executive summary, intervention background, evaluation background, questions and criteria, evaluation methodology, findings, conclusions, recommendations, lessons learned and emerging good practices (see Annex I for more details). Each component was judged in terms of one or more items, each of which in turn was assessed against one or more elements or criteria, to guide the assessment.

Once all elements and items under a particular component were considered, rubrics were used to determine compliance with established criteria and standards, to unpack criteria into standards as shown in Table 6. The next section provides more details.

Finally, the analytical framework included four different scoring systems:

1. Global. The assessment of quality trends in evaluation reports used a comprehensive rating system that combined weighted components with a standardized rubric scale ranging from 0 to 5. This scoring system set five as the maximum achievable score for any evaluation report.
2. Dimension 1. A six-point scale considered frequencies, proportions and arithmetic means. A global rating was assigned to each report, considering the ratings given and the relative weight of each component (see Figure 1).
3. Dimension 2. A dichotomous nominal scale (yes/no categories) with an analysis of frequencies and proportions was used to evaluate the comprehensiveness of evaluation reports (Annex II).
4. Dimension 3. A dichotomous nominal scale (yes/no categories) with an analysis of frequencies and proportions was used to evaluate the comprehensiveness of the ToR (Annex III).

This analysis was combined with a five-level scoring scale developed to assess the completeness and quality of each section. The different rubrics are outlined in Table 6.

Table 6. Overall scoring systems

General (quality trends in evaluation reports)					
0–1.99	2–2.99	3–3.99	4–5		
Misses requirements The report does not meet the minimum requirements of quality considered	Approaches requirements The report meets only some of the key requirements	Meets requirement All the requirements are adequately met	Exceeds requirements The report excels in covering the required elements		
Dimension 1 (quality of evaluation reports)					
0	1	2	3	4	5
Highly unsatisfactory (HU) None of the required	Unsatisfactory (U) Not all the elements are present, and at	Somewhat unsatisfactory (SU) Not all the elements are	Somewhat satisfactory (SS) All the elements are present but	Satisfactory (S) All the elements are present	Highly satisfactory (HS) All elements are present, complete, interconnected and the

elements are present.	least one of the elements presents is incomplete.	present, but those present are complete.	not all of them are complete	and complete.	report excels in covering the item.
<b>Dimension 2 (comprehensiveness of evaluation reports)</b>					
A dichotomous nominal scale (yes/no categories) with an analysis of frequencies and proportions					
<b>Dimension 3 (comprehensiveness of terms of reference)</b>					
A dichotomous nominal scale (yes/no categories) with an analysis of frequencies and proportions					
<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	
Absent	Poor	Moderate	Good	Very good	
The section is missing or does not contain any relevant elements	The section is incomplete or vague, providing minimal relevant content.	Some key elements are present, but the content lacks depth or clarity.	The section covers most elements with sufficient clarity and moderate detail.	The section includes all key elements, clearly defined, with high-quality content.	

To ensure the validity of the evaluation, the team used multiple approaches to triangulate the findings. These included:

- Looking at different types of data from various sources to cross-check information;
- Using multiple research methods to confirm findings;
- Having multiple researchers examine the same information independently;
- Ensuring the measurement tools were accurate by checking whether they captured what they were meant to measure.

In terms of ensuring reliability, several steps were taken to guarantee that data collection methods were consistent and reliable:

- The team members conducted tests of the scoring tools and rating systems during the pilot phase to confirm that these tools were effective and suitable. Each team member assessed the same report independently. This served two key purposes:
  - They compared everyone's scores and explanations to find differences in how team members interpreted the criteria. This helped create a unified scoring approach that everyone understood and followed.
  - Once all evaluations were complete, they double-checked any reports that received particularly high or low scores through peer review.
- The evaluators measured the time needed to assess each evaluation accurately, allowing them to establish how many elements they could realistically analyse and how many evaluation reports they could review within their time frame.

The team also created specific guidelines to ensure that everyone scored consistently across evaluations. These guidelines are available in Annex IV. The inception report in Annex XIV provides more details about the initial testing phase.

Figure 1 shows the weight attributed to each component in the scoring tool, allowing the final overall scoring to be determined.

Figure 1. Relative weight for the rating of each component

Findings 25%	Conclusions 20%	Recommendations 20%	Criteria and questions 8%	Evaluation methodology 8%
			Executive summary 5%	Evaluation background 5%
			Project background 5%	Lessons learned 3%
			Good practices 3%	

#### a. Data sources

The research methodology incorporated two distinct data categories. The first category consisted of findings obtained through the CED analytical framework and quality assessment tool, complemented by key informant interviews and online survey results. The second category encompassed pre-existing documentation that forms the basis for ratings and analysis, as referenced in Table 7.

Table 7. Data sources: Existing documents

List of key documents	Type of information collected
Normative and operational framework. IOM's evaluation policies and guidelines, UNEG reference documents, and toolkits and other related documentation (MOPAN)	Contextualization and deep understanding of key frameworks and strategies considered within IOM.
The evaluation reports, ToRs and inception reports	Key information on which to base the assessment.

#### b. Data analysis approach

The evaluation team applied the average rating as the primary measure to analyse results, ensuring comparability across projects, capturing variations in data, and enabling weighted aggregation when necessary. The average rating was selected because it incorporates all available data points, unlike the median or mode, thereby providing a comprehensive representation of evaluation findings. This approach facilitated benchmarking and trend analysis across multiple projects, regions and time periods, aligning with standard evaluation methodologies used by international organizations to ensure consistency and comparability while enhancing interpretability for stakeholders such as policymakers, donors and implementing partners.

Recognizing the sensitivity of the average rating to outliers, the team implemented specific methods to mitigate their impact, assessing the normality of the data distribution to determine whether the average rating was an appropriate measure. Statistical tests and visual methods, such as histograms and Q-Q plots, were used to identify skewness or deviations from normality, ensuring that the chosen analytical approach accurately reflected the data trends. In cases where extreme values would distort the overall interpretation, strategies such as

outliers' exclusion were applied to assess the robustness of the findings. These adjustments helped preserve the integrity of the results while ensuring that key trends remained clear and interpretable for stakeholders.

### c. Sampling procedures

For the analysis of evaluation reports, an initial universe of 265 evaluations was available for assessment. After excluding those conducted by Artival, 260 evaluations remained. A sampling strategy was implemented to ensure that the sample accurately represented the original distribution, resulting in a sample of 90 evaluations.

In the initial step, four key criteria were chosen to ensure the sample's validity, focusing on the main crucial aspects for evaluating the assessments: year; type of evaluator; evaluation coverage;<sup>12</sup> and regions covered. The objective was to consider these four criteria together rather than in sequence. This approach prevents any individual criterion from overshadowing others, which could lead to the underrepresentation of significant categories in the final sample. Combining these criteria created a measure of each value's weight relative to the total population. These weights were then used as probabilities in a Python-based random sampling algorithm, resulting in the sample of 90 evaluations.

In the second step, two additional criteria – evaluation scope;<sup>13</sup> donor – were selected to verify the balance of the sample. Based on these criteria, the selected sample was compared with the total population to ensure that the most relevant categories related to the type of evaluation scope and the donor were adequately represented in the final sample.

Some categories were omitted during the sampling process. This was anticipated, as the primary goal was to reduce the number of evaluations while maintaining a representative sample. For the “regions covered” criterion, evaluations that involved countries from more than a single region were excluded, and for the “evaluation scope” criterion, synthesis evaluations were likewise excluded. The final sample was considered valid, since individual evaluations containing these categories were included.

As a result, approximately 3 per cent of cases for regions covered and about 6 per cent for evaluation scope were not represented in the final sample. Additionally, some donors were absent from the final sample due to their diversity in criteria. Overall, only 5 per cent of donors were not represented: for example, reports with funding from the Government of the Kingdom of the Netherlands, which had 3 out of the initial 260 evaluations, or from UNICEF and the Government of France, each with only one evaluation, were not included.

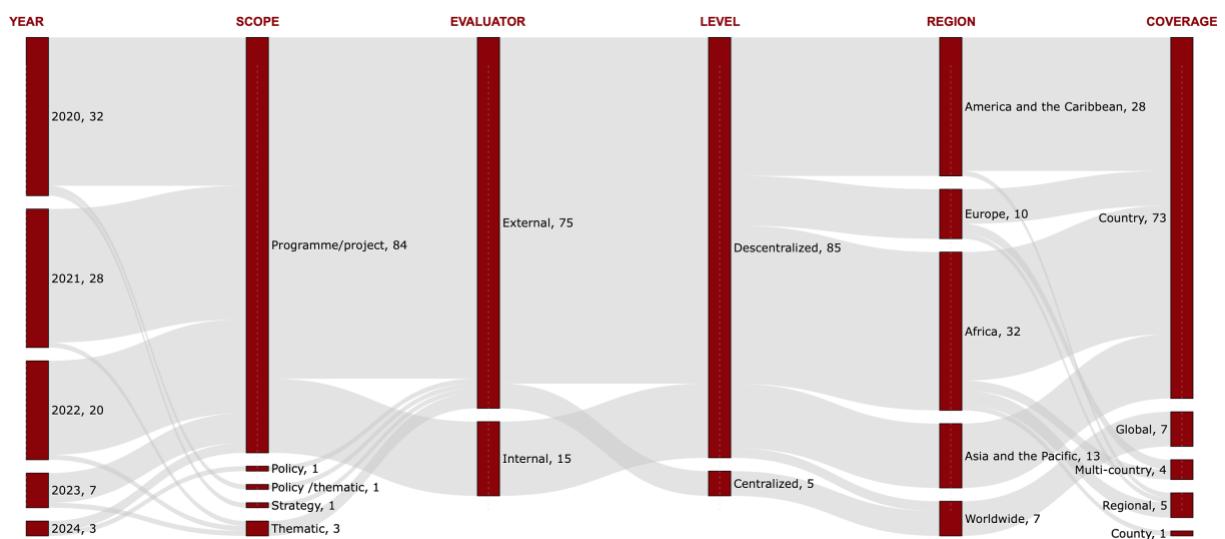
Figure 2 shows the distribution of the evaluations under assessment per the different criteria used: year; scope; type of evaluator; region; coverage.

---

<sup>12</sup> The evaluation coverage includes the following categories: global, regional, multi-country and country evaluations.

<sup>13</sup> Policy, thematic, strategy, synthesis or programme/project evaluation.

Figure 2. Overview of evaluations under assessment



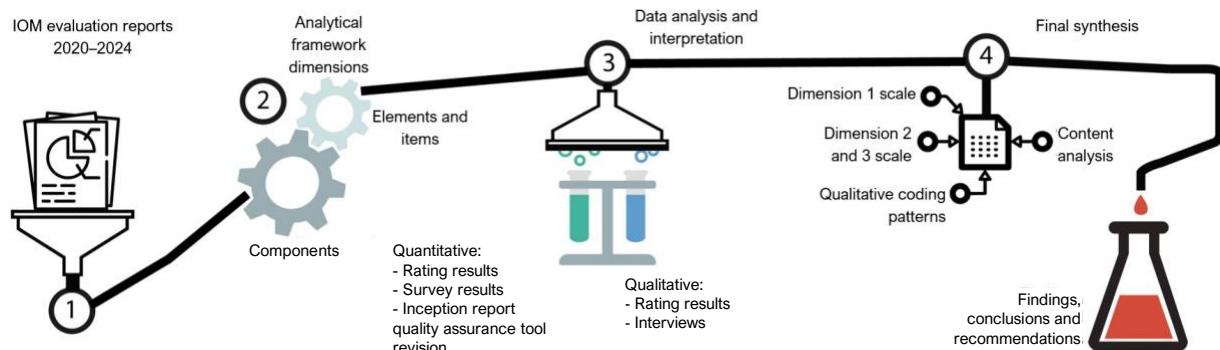
From the 90 selected evaluations, a sample of 20 ToR (10 from the highest-scoring evaluations and 10 from the lowest-scoring evaluations) was analysed qualitatively and quantitatively using the existing quality control tools.

The sampling strategy was applied to all 58 ToRs within the 90 evaluation reports reviewed in the meta-evaluation. For dimension 3, which measures the comprehensiveness of the ToRs, a yes/no nominal scale was used to assess frequencies and proportions. Additionally, a more nuanced assessment of 20 ToRs was conducted, considering the criteria of the IOM checklist for ToRs. The selection included the two highest-scoring, the two lowest-scoring, and one mid-range report for each year from 2020 to 2022. The sample is smaller for 2023 and 2024 due to a reduction in the number of evaluation reports produced during that period. Thus, for 2023, four ToRs were selected: two from the reports with the highest scores; two from the reports with the lowest scores. From 2024, only the single available ToR was assessed. This double selection criterion identified the potential link between the quality of ToRs and the evaluation reports' quality over annual periods, providing both a diachronic and synchronic perspective. One of the 20 ToR was excluded because it was developed alongside the inception report, meaning it did not adhere to the standard ToR structure and could not be assessed independently. This adjustment resulted in a final sample of 19 ToRs. The list of ToR analysed can be consulted in Annex VII.

Additionally, three inception reports were reviewed using the IOM quality control tool for inception reports launched in 2022: one from 2020; one from 2022; one from 2023. These reports were randomly selected, ensuring that the overall score of the evaluation report with which they were linked was not known in advance. The goal was to represent three periods: the first (2020), when a significant number of evaluations were conducted before the publication of the quality assurance tools for inception reports, ToRs and final reports; the second (2022), with a substantial number of evaluations carried out using the available tools; and the third (2023), as the most recent period, given that there was no access to inception reports for 2024. The reports were selected from a pool of 25 inception reports attached to sampled evaluation reports. The inception reports analysed can be consulted in Annex VII.

The following figure illustrates the overall assessment process, from the final set of evaluation reports being considered to the final report's findings, conclusions, and recommendations.

Figure 3. Overall process of the meta-evaluation



The external team from Artival Research & Evaluation performed the assessment, guided by the inception report (Annex XIV) and ToR (Annex XIII). Throughout this process, IOM examined deliverables, contributed input and granted final report approval. Before finalization, IOM conducted a comprehensive review of all deliverables, offered detailed feedback and formally approved the final report.

Survey questionnaires targeted three distinct stakeholder groups: external evaluators (382), internal evaluators (17) and evaluation managers (420). The targeted internal and external evaluators included those who have conducted evaluations published in the IOM repository, and those who potentially have conducted evaluations, even if they are not included in the repository. The evaluation manager survey was sent to 420 staff members who held management responsibilities for projects that underwent evaluation during the analysed period. In all three survey, it was requested that only those involved in one of the respective roles during the period respond to the survey. The internal evaluation survey received 24 responses. After excluding incomplete responses, 17 were retained for the analysis. Similarly, the external evaluation survey received 66 responses, 50 complete, and the evaluation managers survey 109, out of which 36 were complete and used for the analysis.

#### d. Cross-cutting issues

Gender, human rights, disability, and environmental integration were the cross-cutting issues that the meta-evaluation looked at. The assessment determined how a human rights-based approach, gender equality, disability inclusion and environmental sustainability were considered when designing evaluations and defining findings, conclusions and recommendations.

The evaluation methodology incorporated dedicated criteria to assess these cross-cutting issues at three critical stages. First, in the analysis of evaluation methods, one of the criteria examined how thoroughly cross-cutting issues were integrated into the evaluation design and execution. Second, in the analysis of the findings of each evaluation, one criteria analysed the extent to which these issues emerged from and influenced the evaluation results. And third, in the analysis of each evaluation's recommendations, one of the criteria evaluated whether the proposed actions adequately address and account for relevant cross-cutting

considerations. The evaluation approach and methods of data collection and analysis were analysed to determine whether they were rights-based and gender-responsive and whether data was disaggregated by social criteria (such as gender, ethnicity, age, disability, geographic location, income and education).

In assessing the comprehensiveness of evaluation reports, a detailed breakdown was produced examining each cross-cutting issue separately. This granular analysis allowed a more precise assessment of how individual cross-cutting approaches were incorporated into the evaluation reports. The enhanced structure enabled evaluators to systematically track and measure the depth of inclusion for each specific cross-cutting element, rather than treating them as a single consolidated category.

#### e. Norms and standards

The meta-evaluation adhered strictly to the UNEG foundational ethical and professional conduct principles. The consulting team operated within the established framework of relevant IOM policies and UNEG norms and standards,<sup>14</sup> maintaining objectivity throughout the assessment process. Their evaluation demonstrated integrity and fairness, supported by formal declarations confirming the absence of conflicts of interest. Specifically, the consultants verified they had neither direct nor indirect involvement in the evaluated interventions nor personal connections with the management teams or consulting personnel associated with these assessments.

#### 2.5. Limitations of the meta-evaluation

The assessment team faced the following key limitations during the meta-evaluation process. Other potential limitations were considered during the initial phase but were not ultimately faced during the analysis.

Table 8. Risks, limitations and mitigation strategies

Risk or limitation	Explanation	Mitigating strategy
Limited number of evaluations with information on budget assigned and training received by internal evaluator	Two of the meta-evaluation's key questions aimed to identify whether differences in the quality of evaluation reports related to the budget utilized or the training received by the internal evaluator.	The team included an analysis of the results of those evaluations with information on costs and training of internal evaluators. The relevance of the analysis was based on the representativeness of the results.
Different types of evaluations	There were various types of IOM evaluations, including internal or external assessments, as well as mid-term, final, or ex-post evaluations. Additionally, some evaluations focused on specific programmes or projects, themes, or strategic aspects. Significant differences may hamper the use of common assessment criteria across the different evaluations assessed in the meta-evaluation.	The meta-evaluation considered this and decided to discard synthesis reviews. This was done to ensure consistent application of the scoring tool, as the synthesis reviews did not align with the exercise's logic.
Lack of information on training and evaluation budgets	The absence of systematically recorded data on training received by evaluators and on budgets allocated to evaluations made it difficult to analyse the relationship between evaluator training, budget allocated in evaluations, and the quality of the evaluations. Additionally, the variability in how this	Institutional reports, internal documents and organizational databases were reviewed by the evaluation manager to gather relevant data on training and budgets. Additionally, specific questions were incorporated into interviews and surveys to obtain qualitative insights that could

<sup>14</sup> UNEG guidance, practices and norms can be accessed from [this web page](#). See especially UNEG, *Norms and Standards for Evaluation* (New York, 2016).

Risk or limitation	Explanation	Mitigating strategy
	information was recorded and reported hindered consistent comparisons across different cases, potentially affecting the robustness of the conclusions.	supplement the available data. This triangulation of sources enabled a more robust analysis, minimizing the impact of the lack of systematized information and providing a stronger foundation for assessing the relationship between evaluator training, evaluation budget, and the quality of the evaluations.
Imbalance of findings per year, coverage and region	The evaluation sample presents a significant fragmentation that affects the comprehensiveness and representativeness of the meta-evaluation findings. The sharp decline in the number of evaluations since 2023, the regional disparities with some being overrepresented, and the uneven coverage of thematic areas increase the risk of bias and reduce the robustness of trend analysis.	Findings were carefully contextualized by year, and regional and thematic disparities were mitigated by disaggregating the analysis by region and thematic areas.
Interviews limitations	Some of the evaluations were carried out by external evaluators, and others by internal staff who were no longer in the Organization, so it was not possible to access them.	The consulting team conducted a purposive sampling focusing on those informants who were reachable, to collect relevant data for the analysis.
Survey participation	The low response <sup>15</sup> rate to the survey sent to evaluators and evaluation managers affected the representativeness of the findings, reducing the ability to generalize results and limiting the diversity of perspectives considered in the analysis.	The survey was disseminated by IOM and periodic reminders were sent to encourage participation. Data collected were complemented with qualitative information from interviews and document reviews, allowing for triangulation of findings and strengthening the validity of the results.
Inclusion of 2024 data with a limited sample	The inclusion of 2024 in the analysis presented a limitation due to the low number of evaluations available in the sample (three out of a total of eight) by the time data was analysed. This limited representation meant that the 2024 data only reflected an emerging trend, without allowing for the identification of consolidated patterns or long-term tendencies. As a result, any variation observed in this year should be interpreted with caution, as it may not be representative of the overall behaviour of evaluations in previous periods.	The 2024 data have been treated as relative within the analysis, emphasizing their preliminary nature. Rather than drawing definitive conclusions about trends in this year, the analysis has focused on a comparative process across different years in the sample. This approach helped contextualize recent data in relation to previous periods and highlights potential continuities or changes without overinterpreting the emerging trends of 2024.

## Findings on evaluation report quality

This section examines the quality trends in evaluation reports from 2020 to 2024, using the meta-evaluation scoring tool as the primary assessment framework. The analysis captures recent developments and draws insights from comparison with the results of the 2017–2019 meta-evaluation, providing a comprehensive view of quality evolution over time. The findings presented reflect a systematic assessment of report quality across these time frames, enabling the identification of long-term patterns and shifts in evaluation standards. The meta-evaluation questions (see Table 4) addressed in this section are 1, 4, 5 and 6.

<sup>15</sup> The survey response rates were as follows. External evaluators: 67 responses received, 50 complete and used for analysis. Internal evaluators: 25 responses, 17 complete and used for analysis. Evaluation managers: 39 responses, 34 complete and used for analysis.

## **Key findings**

1. The overall assessment for the 2020–2024 period indicates that 39 per cent of the reports meet or exceed all requirements, 44 per cent approach requirements, and 17 per cent fail to meet the minimum quality standards. The global arithmetic average of evaluations is rated as “approaches requirements,” with a score of 2.8 out of 5. This marks an improvement of 0.2 points from the previous assessment of 2017–2019, which had an average score of 2.6 out of 5.
2. The trend from 2020 started at a high of “meets requirements” with a score of 3 out of 5. It then dropped to 2.6 in 2022 but began recovering in 2023, reaching “approaches requirements” at 2.7. The average score for the 2017–2019 assessment was 2.4 in 2017, remaining at 2.7 for both 2018 and 2019, consistently staying under the “approaches requirements” level. This indicates that while the 2020–2024 period experienced fluctuations, it started at a higher level and showed signs of recovery, whereas the 2017–2019 period maintained a lower, more stable performance.
3. At the regional level, reports from America and the Caribbean and from Europe score above the average (2.9). In contrast, reports from Asia and the Pacific and from Africa score slightly lower. Compared to 2017–2019, scores for reports from the global, America and the Caribbean and Africa regions improved, while scores for reports from Europe and Asia and the Pacific declined.
4. There is no major quality difference between reports by internal and external evaluators. Internal evaluators kept the same average score as in the previous meta-evaluation, while external evaluators improved their score from 2.6 to 2.8 during 2020–2024. On average, evaluations still rate at “approaches requirements”.
5. Evaluations in the thematic areas of media and communications, assisted voluntary return, counter-trafficking, multiple thematic and migration health achieve the “meets requirements” rating, on average.
6. The global scoring results by geographical scope reveal that evaluations conducted at a multi-country level (80% of the total sample) achieve the highest scores, followed by those with a global scope that also (though barely) rate at “meets requirements”. When assessing results by modality, strategy and policy evaluations lead with a score of 3.3, followed by programmes and projects, which score 2.8. In terms of timing, ex-post evaluations receive the highest scores.
7. The analysis suggests that the absolute size of an evaluation budget has little to no meaningful impact on evaluation ratings. However, the proportion of the total budget allocated to evaluation shows a stronger and statistically significant relationship with evaluation quality.
8. The sections of evaluation reports that achieved a score higher than the global average (2.8) are findings (3.3), and evaluation background, criteria and questions and conclusions (3.1), all of which reached a rating of “meets requirements”. Sections that scored at the

“approaches requirements” level are project background (2.7), evaluation methodology and recommendations (2.5), and executive summary (2.3). The sections that scored at the “misses requirements” level are lessons learned (1.6) and good practices (0.8).

9. Comparing the sections of reports to those assessed in the 2017–2019 meta-evaluation shows that the project background and conclusions sections have notably decreased in quality, by 0.6 and 0.9 points, respectively. In contrast, the scores for criteria and questions, evaluation methodology, and findings have improved significantly, with criteria and questions experiencing the largest increase of 0.8 points. Other components of evaluation reports that improved their score compared to the previous meta-evaluation are the evaluation brief, list of tables, figures, and charts, list of acronyms or abbreviations, executive summary and lessons learned. The rest remained relatively unchanged. There is still a wide margin for improvement for the cross-cutting issues of disability inclusion (included in only 10 per cent of reports) and environmental sustainability (8%), as well as the good practices section (included in only 32% of the evaluations assessed).

10. A positive correlation has been identified between the quality of the ToRs and the quality of evaluation reports. Weak sections on cross-cutting issues, on methodology, and on the formulation of lessons learned and good practices can negatively impact the evaluation’s development and outcomes.

11. The analysis of inception reports illustrates the importance of a well-designed evaluation matrix as a guiding tool in the evaluative process. Similarly, a thorough analysis of the adequacy of the methodology is crucial for improving evaluation quality.

### 3.1. Overall findings on evaluation report quality

Between 2020 and 2024, the analysis of 90 evaluation reports (see Figure 4) shows that 39 per cent (35 reports) met or exceeded requirements. Another 45 per cent (40 reports) scored at the “approaches requirements” level. Meanwhile, 17 per cent of the reports failed to meet minimum quality standards (15 reports). The overall rating for this period averaged 2.8 points, a 0.2-point increase compared to the previous three-year period (2017–2019). This means that, on average, evaluation reports approach requirements, and are very close (0.2 points away) from achieving “meets requirements”.<sup>16</sup>

The analysis of reports from 2020 to 2024 reveals a trend of declining quality, reaching its lowest point in 2022 before beginning a recovery in 2023. Specifically, while over 90 per cent of the reports were rated positively in 2020, this percentage dropped to 63 per cent by 2022, with 37 per cent of evaluations deemed below the minimum standard.

In 2023, evaluation reports improved, with 43 per cent marked as “meets requirements” and another 43% marked as “approaches requirements” (43%), but none marked as “exceeds requirements” (see Figure 4).

<sup>16</sup> In assessing the quality of the evaluation reports, each component was assigned a specific weight that contributed to the overall assessment of the evaluation report. For example, the quality of the evaluation report’s recommendations contributed more significantly to the overall rating of the report than did the quality of the executive summary.

Figure 4. Evolution of global rating 2017–2024 per levels

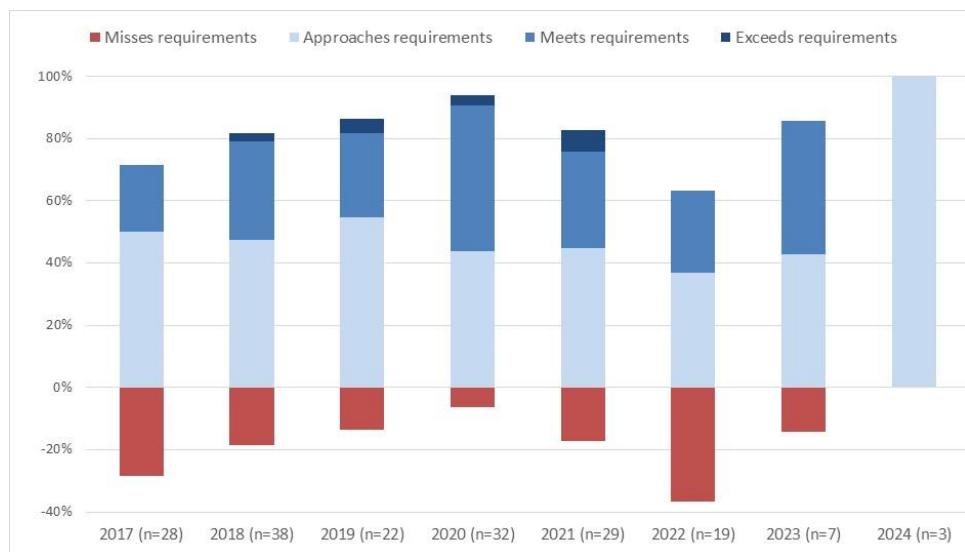
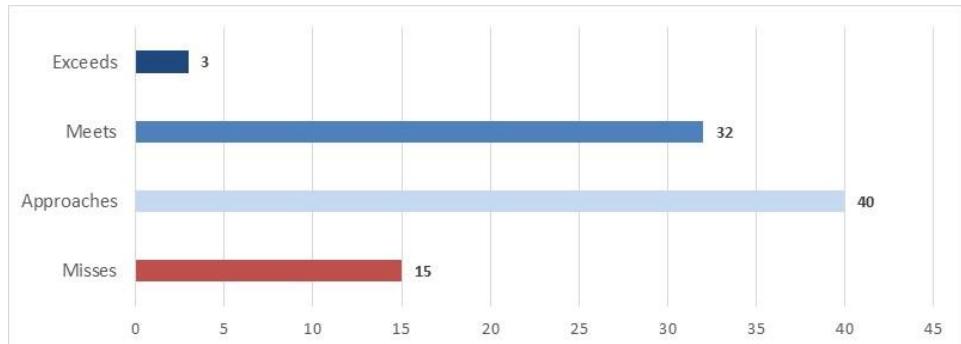
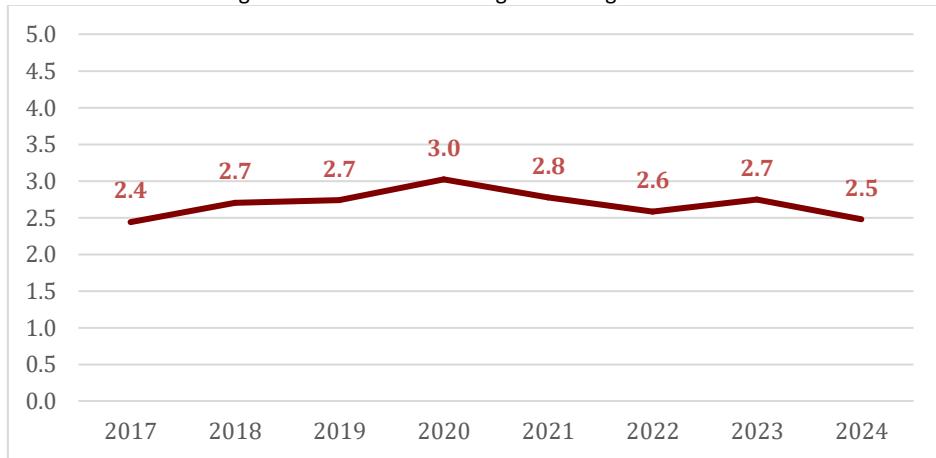


Figure 5. Distribution of reports per rating 2020–2024



The mean overall rating evolution per year shows a positive trend from 2017 to 2020, increasing from “approaches requirements” (2.4) to “meets requirements” (3). However, a decline began thereafter, reaching 2.6 in 2022. A slight recovery occurred in 2023, with the rating rising to 2.7, which remains at the “approaches requirements” level (see Figure 6). This recovery restored the rating to its 2019 level (2.7), although no evaluation report exceeded requirements during this period.

Figure 6. Evolution of mean global rating 2017–2024



The survey also included questions related to quality improvement. When evaluation managers were asked about the key factors influencing the quality of evaluation reports, over half (53%) identified stakeholder engagement and participation during the evaluation process as the most significant elements. About 17 per cent highlighted IOM quality assurance tools and templates as key to improving evaluation quality. Meanwhile, 14 per cent cited aligning findings with organizational priorities, and another 14 per cent cite clear evaluator recommendations, as important. Only 3 per cent saw the timing of the evaluation in the programme cycle as a major factor.

Key informants consulted during the interview phase consistently emphasized the evaluator's expertise, the robustness of the methodology and the strength of the quality assurance process as the most decisive elements in determining the quality of the evaluation. The selection of evaluators played a crucial role, as reports often reflect the depth of knowledge, analytical skills and methodological rigour of the person leading the evaluation. However, inadequate selection criteria can sometimes result in consultants being chosen based on availability rather than technical competence. The methodological framework was also recognized as a significant factor influencing quality, as evaluations that lack a clear sampling strategy, data triangulation, or alignment between findings and recommendations often produce inconsistent or overly descriptive reports rather than analytical ones. Another critical factor was the level of oversight and feedback throughout the evaluation process. While regional offices and evaluation managers provide guidance in some cases, their involvement varies, and in many instances reports are finalized without systematic quality checks. Institutional tools such as standardized templates and quality assurance checklists have enhanced structure and consistency. Still, their use remains inconsistent, particularly in country offices with limited monitoring and evaluation capacity.

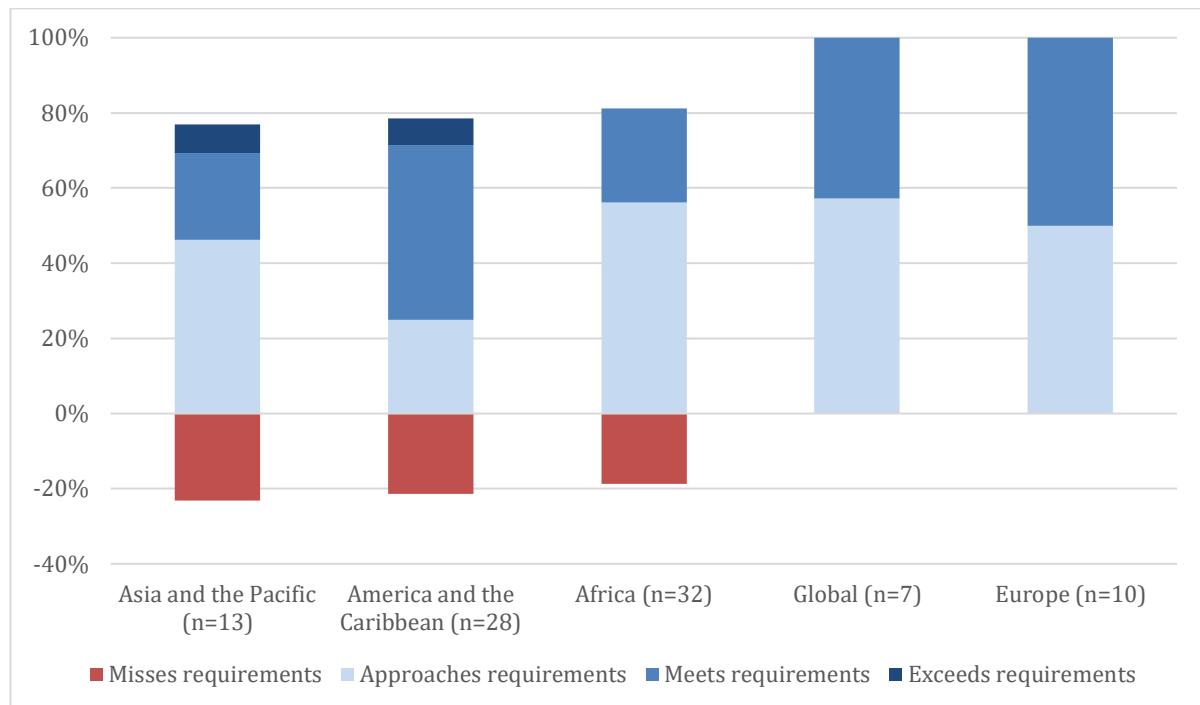
### 3.2. Geographical analysis

The distribution of results by region shows that evaluations with a global scope and those with a geographical focus in Europe either approach or meet requirements (see Figure 7).<sup>17</sup> In Europe, 50 per cent of evaluations are rated "meets requirements" and 50 per cent "approaches requirements," whereas in the global category a higher proportion (57%) of evaluations rated "approaches requirements" and 43 per cent rated "meets requirements".

---

<sup>17</sup> The geographical analysis aligns with the one from the 2017–2019 meta-evaluation, which differs to some extent from the coverage of IOM regional offices. A detailed breakdown is available in Annex VI. For non-global evaluations covering countries in more than one region, only one region is displayed.

Figure 7. Global rating levels per region

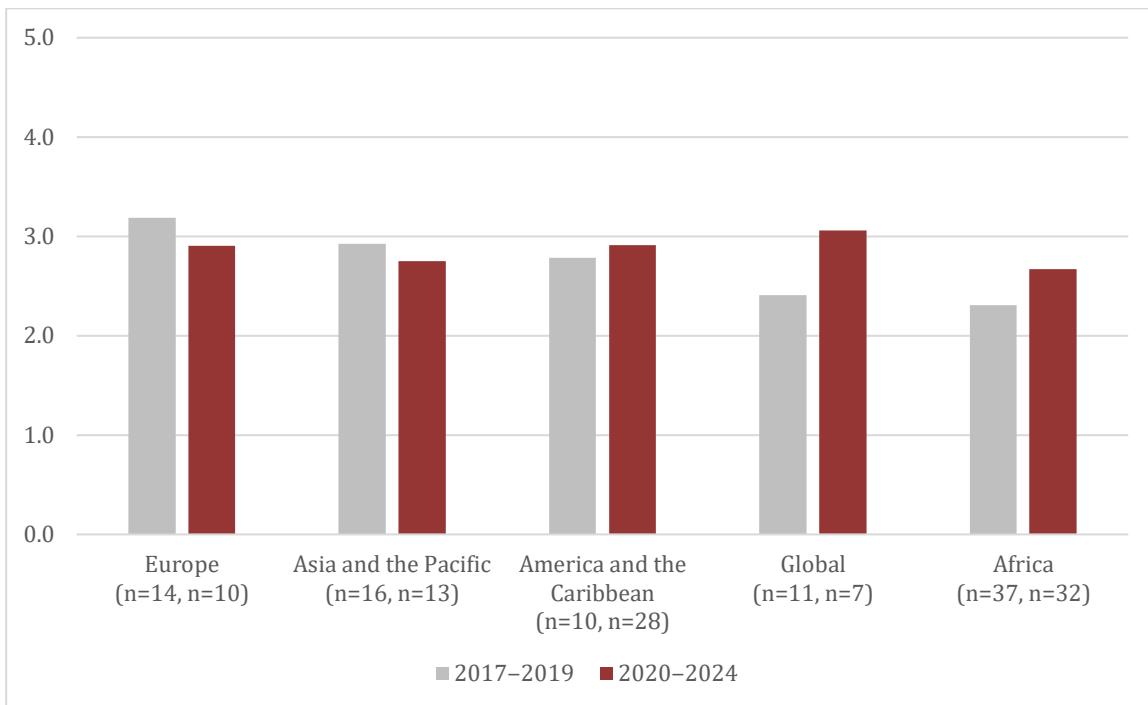


The Asia and the Pacific region and the America and the Caribbean region exhibit the most polarized results. In the Asia and the Pacific region, 8 per cent of reports are classified as “exceeds requirements,” while 23 per cent are categorized as “misses requirements.” Similarly, in the America and the Caribbean region, 7 per cent of reports are categorized as “exceeds requirements” and 21 per cent as “misses requirements”. In contrast, 19% of reports from the Africa region are categorized as “misses requirements”, 56 per cent are classified as “approaches requirements” and none are categorized as “exceeds requirements”.

Despite having 21 per cent of evaluations that did not meet the minimum requirements, the America and the Caribbean region achieves an average rating of “approaches requirements” (2.9 out of 5), thanks to the number of evaluations that exceed the requirements. Meanwhile, the Asia and the Pacific region and the Africa region receive the lowest average rating of 2.7 out of 5. Nevertheless, both regions are also at the “approaches requirements” level.

The regional average rating adds valuable insights to the previous data (see Figure 8). The Europe region shows average levels that range between approaching and meeting the requirements, but its average rating is slightly below 3, at 2.9 out of 5. On the other hand, the global region’s average rating surpasses this, scoring 3.1 out of 5. This makes the global region the only one that qualifies as “meets requirements.”

Figure 8. Global mean ratings per region (2017–2019 and 2020–2024)



The comparison with the average rating for the 2017–2019 period shows that the global, America and the Caribbean, and Africa regions have improved their results, while the Europe and Asia and the Pacific regions have worsened them.

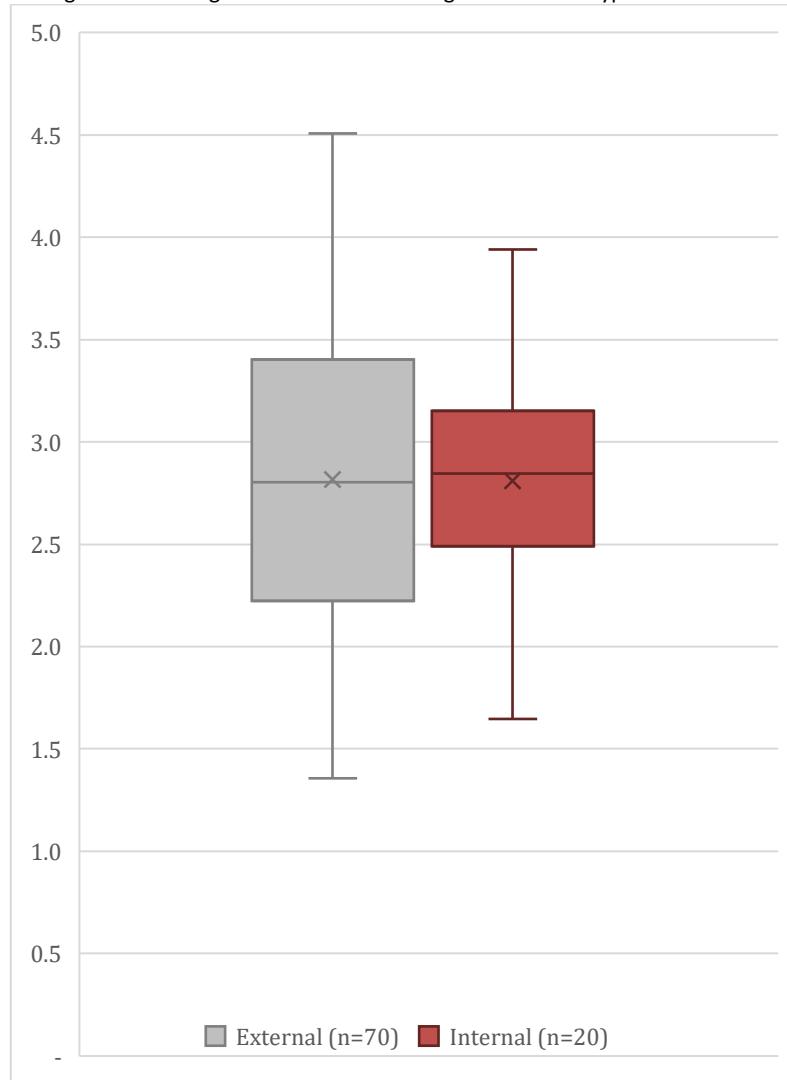
### 3.3. Analysis by evaluator

The assessments of evaluation report quality, when comparing those conducted by internal evaluators with those conducted by external evaluators, showed a similar average rating (2.8). This places both groups at the top of the “approaches requirements” level. Compared to the 2017–2019 period, internal evaluators have maintained the same average rating, while external evaluators have increased the rating by two-tenths, from 2.6 to 2.8. Figure 9 shows that external evaluators have a broader score range (1.4 to 4.5) compared to internal evaluators, whose scores range from 1.6 to 3.9.

For both evaluations conducted by internal evaluators and for those conducted by external evaluators, 84% of the ratings span between “approaches requirements” (2.0–2.9) and “meets requirements” (3.0–3.9). However, the data from internal evaluators show a narrower spread due to a significantly smaller sample size (20 reports evaluated) compared to external evaluators (70 out of 90 total evaluations).

Of the internal evaluators sampled, 13 out of 15 have received training in evaluation from IOM. Because so few untrained evaluators were assessed, it is impossible to establish a valid relationship between the training received and the quality of evaluations. With only two evaluators untrained, any differences noted in their evaluations compared to those conducted by trained evaluators may be attributed to chance or individual factors such as prior experience, analytical skills, or understanding of the evaluated context. Methodologically, such a small sample lacks the diversity and variability necessary for drawing generalizable conclusions, indicating that findings cannot be applied to other groups or similar situations.

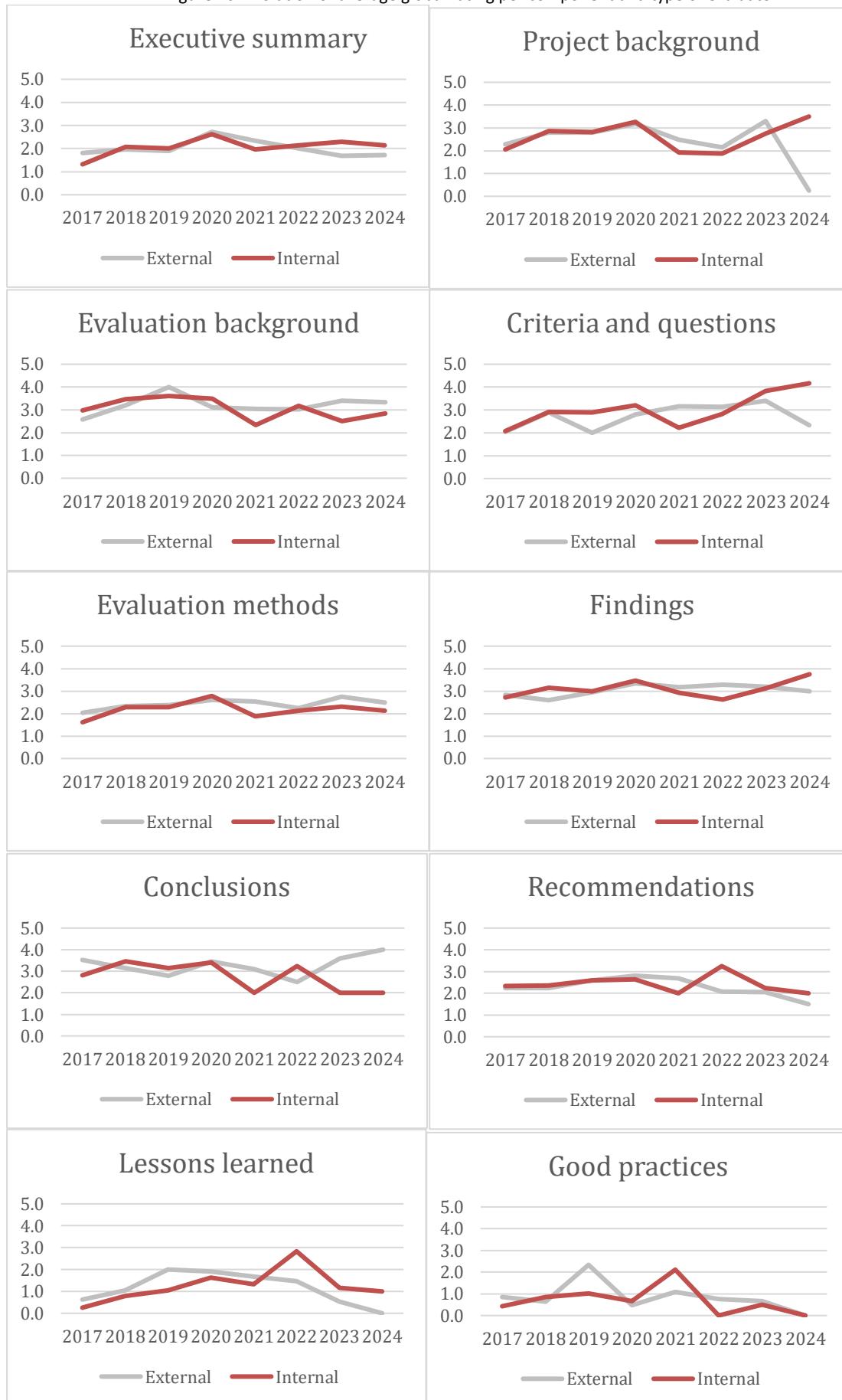
Figure 9. Box diagram of evaluation rating based on the type of evaluator



The analysis by evaluator type and evaluation components shows that the score differences are not significant. However, specific trends emerge when examining individual sections:

- Evaluations produced by external evaluators score higher in evaluation background, methodology and conclusions.
- Evaluations performed by internal evaluators outperform those produced by external evaluators in recommendations, lessons learned and good practices.
- For executive summary, project background, criteria and questions and findings sections of evaluations, a notable shift occurred in 2023, with results increasingly favouring evaluations produced by internal evaluators since that year.

Figure 10. Evolution of average global rating per component and type of evaluator



Surveys revealed varying perceptions of evaluation quality standards based on respondents' roles:

- Evaluation managers generally thought that internal team communication was fluent and that the criteria and questions used for the evaluations were adequate, but often struggled to find qualified internal evaluators.
- External evaluators were satisfied with the methodological aspects but raised concerns about evaluation timelines and the alignment of assessment stages with project cycles.
- Internal evaluators echoed these concerns, particularly disagreeing with workplans that failed to properly coordinate desk reviews, fieldwork and final report presentations within set time frames.

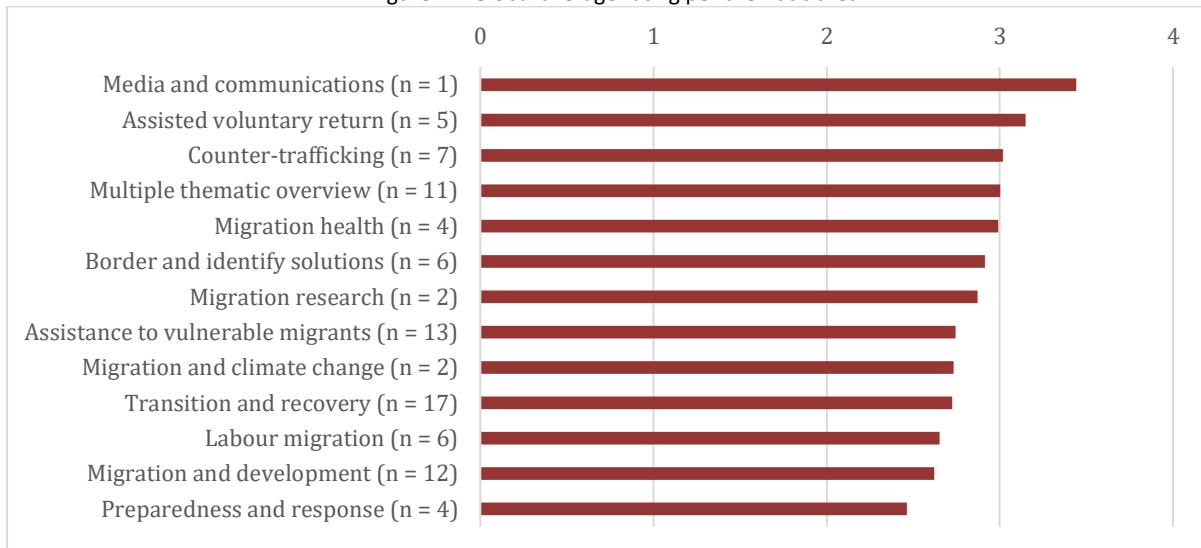
For evaluation managers, the most recurrent challenge was the identification of the evaluators, while for internal evaluators, it was inadequate planning or unrealistic workplans, and for external evaluators, it was limited access to or response from external stakeholders. However, in many instances, the challenges faced by internal and external evaluators were the same.

From a more general perspective, the surveys also revealed different perceptions regarding the quality of the evaluations, depending on the respondent's role. First, as already noted, evaluation managers were most likely to feel positive about internal communication and the adequacy of criteria and questions used. At the same time, they highlighted frequent difficulties in identifying available and adequate internal evaluators. Second, external evaluators were generally satisfied that the evaluation methodology adopted was adequate for the expressed purpose. Still, they expressed more concern regarding the time frame of the evaluations and the stage of the project cycle in which the evaluation was launched and conducted. Finally, internal evaluators shared similar concerns, in general dissatisfied with the workplan and with the time frame that covers and connects the desk review, field phase and final report presentation.

#### 3.4. Analysis by thematic area

The evaluation reports analysed were categorized into 13 thematic areas. While these predefined categories were used in the evaluation repository to classify the evaluations (see Figure 11), it is important to note that most interventions address multiple topics simultaneously, as reflected in the evaluations.

Figure 11. Global average rating per thematic area



Evaluation reports addressing the following thematic areas met or exceeded requirements. The highest score was in “media and communications”, though based on only one report.<sup>18</sup> Following this were “assisted voluntary return”<sup>19</sup>, “counter-trafficking”, “multiple thematic overview”, and “migration health”, which collectively amounted to 27 reports and accounted for 31 per cent of the overall total.

The evaluation reports covering the other eight thematic areas were rated at the “approaches requirement” level. “Border and identity solutions” received the highest score in this group, while “preparedness and response” received the lowest. Reports addressing these eight areas account for 69 per cent of the total. No area falls into the lowest, “misses requirements”, category.

Table 9. Evolution of global rating per thematic area

	2020	2021	2022	2023	2024
Media and communications	n/a	n/a	3.4	n/a	n/a
Assisted voluntary return	3.4	3.3	2.9	3.3	n/a
Counter-trafficking	3.3	3.0	1.9	n/a	n/a
Multiple thematic overview	3.6	2.8	n/a	n/a	2.6
Migration health	3.5	2.4	n/a	n/a	n/a
Border and identify solutions	2.7	n/a	3.3	2.6	n/a
Migration research	3.2	2.5	n/a	n/a	n/a
Assistance to vulnerable migrants	3.7	2.8	2.4	3.1	n/a
Migration and climate change	n/a	2.7	n/a	n/a	n/a
Transition and recovery	2.9	2.8	2.6	2.0	n/a
Labour migration	2.5	n/a	n/a	3.2	n/a
Migration and development	3.0	2.6	1.9	n/a	n/a
Preparedness and response	1.4	3.2	3.4	1.8	n/a

In the period from 2020 to 2024 (see Table 9), the quality trend by topic areas demonstrates a downward trajectory for “counter-trafficking”, “multiple thematic overview”, “transition and recovery”, and “migration and development”, with the latter area experiencing the most significant decline (1.1 points). “Preparedness and response” maintained an upward trend but

<sup>18</sup> IOM, “Evaluación final del proyecto ‘promoción de contribuciones positivas de la migración en América del Sur a través de los medios de comunicación y su interacción con OIM’” (31 January 2022).

<sup>19</sup> Including projects providing reintegration assistance.

saw a decrease again in 2023. Only “assisted voluntary return and reintegration” and “assistance to vulnerable migrants” slightly recovered in 2023, yet they remained 0.1 and 0.6 below their 2020 levels, respectively.

Comparing thematic areas with the previous meta-evaluation presents several challenges. First, there is no direct overlap between the 16 thematic areas assessed in the previous meta-evaluation and the 13 areas analysed for 2020–2024. As a result, only 12 areas could be compared, with “border and identity solutions” excluded due to insufficient data. Second, some areas have data available for only one or two years, offering a limited snapshot rather than a complete trend over time. Despite these limitations, the comparison still provides valuable insights into the evolution of thematic areas in evaluations.

Figure 12. Comparison of global ratings per thematic area and periods.

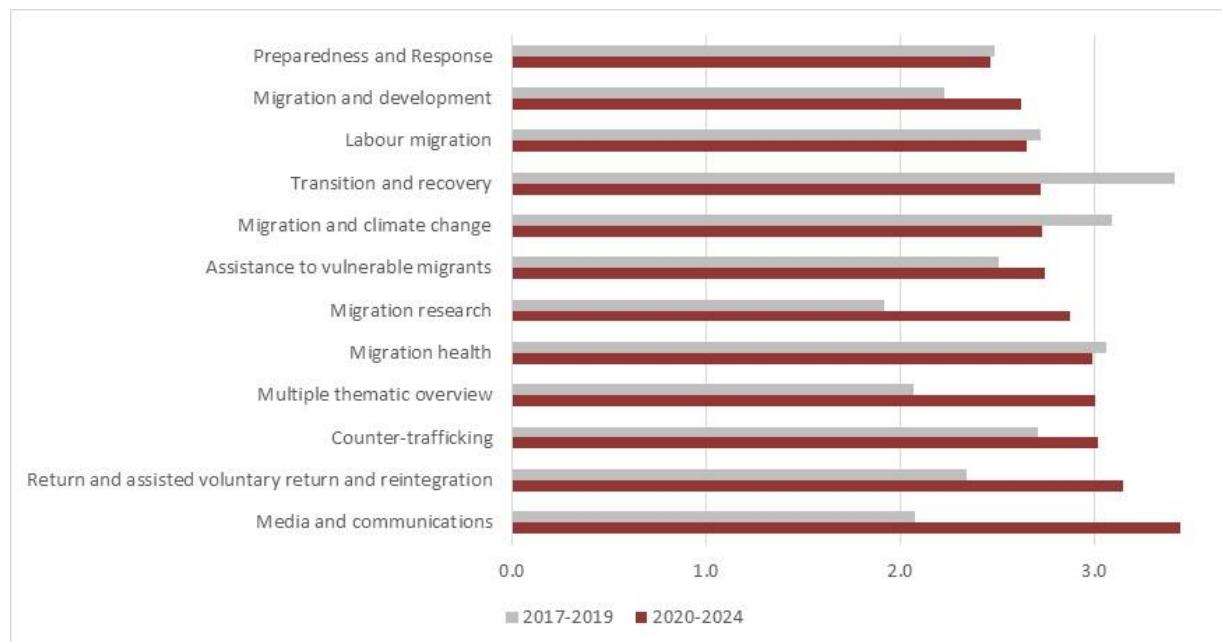


Figure 12 shows global improvements across nine topic areas. “Media and communications” stand out with a 1.5 point increase, though only one report was assessed in each period. “Assisted voluntary return and reintegration” follows with a 0.8 point increase. Three areas show a negative trend, notably “transition and recovery”, which declined by 0.7 points from the previous meta-evaluation. Only “migration health” maintained its overall score.

### 3.5. Analysis per evaluation scope

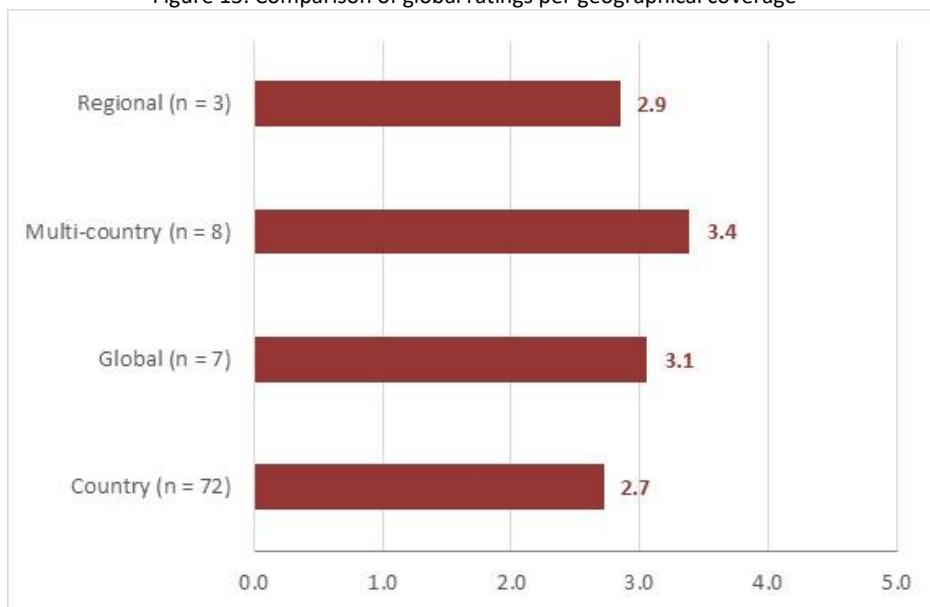
Regarding geographical coverage, the evaluation reports were categorized into country, multi-country, regional and global. Most of the evaluations were conducted at the country level (80%) followed by the multi-country (9%), global (8%), and regional levels (3%).

The analysis of global scores reveals that evaluations conducted across multiple countries achieved the highest ratings, exceeding the “approaches requirements” average (2.8) by 0.6 points. Global-scope evaluations followed closely, scoring 0.3 points above the average at “meets requirements”. Multi-country evaluations have consistently maintained top performance since the 2017–2019 period, sustaining a “meets requirements” score (see Figure

13). Regional evaluations showed consistent average scores, while country-level assessments lagged slightly behind by 0.1 points on average.

The evolution over time was well documented only in country-level evaluations, since the other categories only include one report per year. In this context, the same trend observed at the global level applies: 2020 and 2024 both achieved the highest score (2.9). The most significant decline occurred in 2023, resulting in a score of 2.4.

Figure 13. Comparison of global ratings per geographical coverage



The assessment by modality shows that most reports (91%) related to projects and programmes, while the remaining 8 per cent were distributed among strategic and policy evaluations, thematic evaluations, and one uncategorized report.

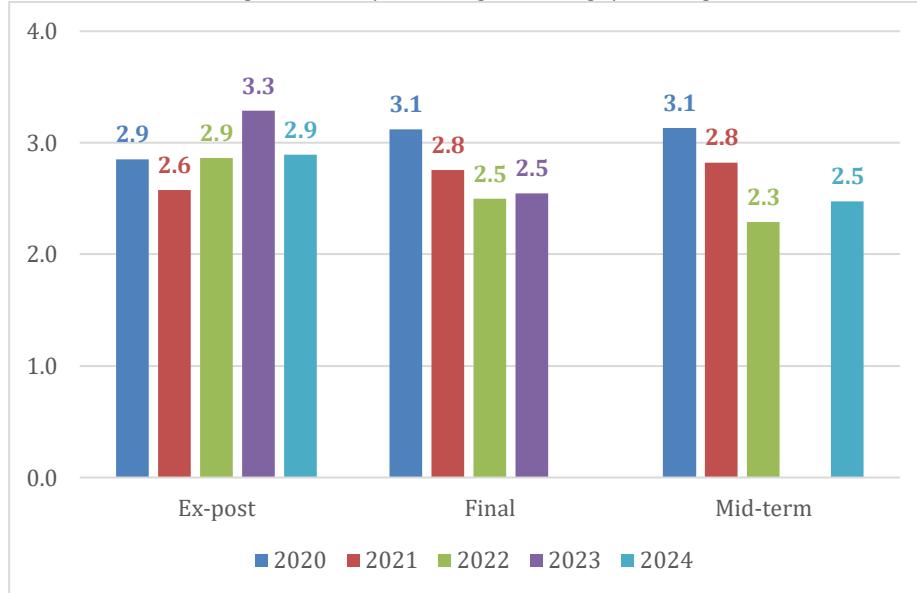
The programme and project evaluations maintained an average rating of 2.8, which coincides with the overall average and is at the high level of “approaches requirements” (see Figure 14). Thematic evaluations exceeded the average, achieving a score of “meets requirements” (3.3), and the strategy and policy evaluations received the lowest score of “approaches requirements” (2.6).

Figure 14. Comparison of global ratings by modality



The analysis of results based on the timing of the evaluation considered the categories of mid-term, final and ex-post evaluations. Among these three categories, final evaluations were the most numerous (52 reports), followed by ex-post evaluations (21) and mid-term evaluations (10).

Figure 15. Comparison of global ratings per timing



Final evaluations peaked at an average score of 3.1 in 2020, but by 2023, this had declined to 2.5, marking a 0.6-point drop. The global average score for final evaluations was 2.8 (“approaches requirements”). Ex-post evaluations conducted since 2022 (in most cases commissioned by the IOM Development Fund) outperformed mid-term and final evaluations. Mid-term evaluations initially averaged 3.1 (“meets requirements”) in 2020 but dropped to 2.3 (“approaches requirements”) in 2022, making them the lowest-scoring category among the three. Their global average score was 2.8 (“approaches requirements”) (see Figure 15).

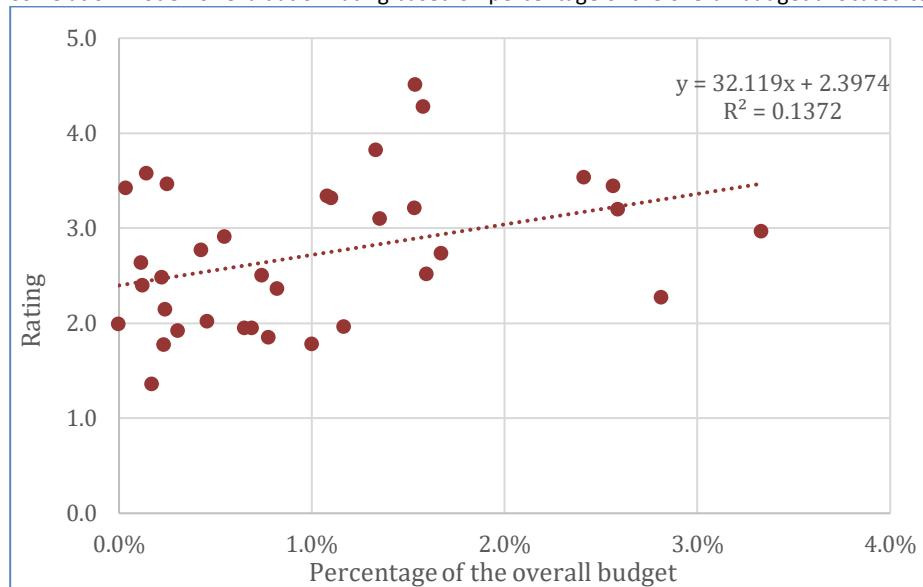
During the interviews, the evaluation managers and the evaluators noted that time constraints were a recurring challenge. Tight deadlines forced evaluators to condense essential steps – including stakeholder consultations, document reviews and validation discussions – into very short time frames, which consequently diminished the overall quality of findings. Rather than serving as strategic learning tools, evaluations often became procedural exercises aimed at meeting the requirements agreed between IOM and the donor instead of providing evidence-based insights aligned with the needs of the project, the country office and key institutional partners.

### 3.6. Analysis per evaluation budget

The evaluation team conducted a correlation analysis to explore whether evaluation ratings were linked to budget size.<sup>20</sup> The first model examined the connection between evaluation ratings and the total evaluation budget. However, the results were inconclusive, as the budget explained only 0.7 per cent of the differences in ratings.<sup>21</sup>

The second analysis examined whether the percentage of the total budget allocated to evaluation influenced evaluation ratings. The results showed a stronger connection than the first analysis, with 14 per cent of the differences in ratings explained by this factor.<sup>22</sup>

Figure 16. Correlation model for evaluation rating based on percentage of the overall budget allocated to evaluation



The survey revealed differing perceptions among evaluation managers, internal evaluators and external evaluators regarding the relationship between financial resources and the quality of evaluation outputs. Forty-two per cent of evaluation managers considered these aspects to be

<sup>20</sup> As noted in the limitations section, the analysis relied on very limited and unverified information. The information was extracted from evaluations conducted by external evaluators, from the general project budget and from the evaluation expenditures in the enterprise resource planning system, rather than an evaluation-specific database. Therefore, conclusions should be interpreted with caution and taken as only partial.

<sup>21</sup> ( $R^2 = 0.0072$ ). The p-value for this model was  $8.44 \times 10^{-11}$ , which means the result was statistically significant (indicating that it is unlikely to be due to chance). However, despite this statistical significance, the relationship was extremely weak, meaning the evaluation budget had almost no practical influence on the ratings.

<sup>22</sup> ( $R^2 = 0.1372$ ). The statistical test confirmed that this relationship was highly significant ( $p\text{-value} = 2.29 \times 10^{-21}$ ), meaning it is very unlikely to be due to chance. This suggests that when a larger share of the budget is dedicated to evaluation, the quality of the evaluation tends to improve.

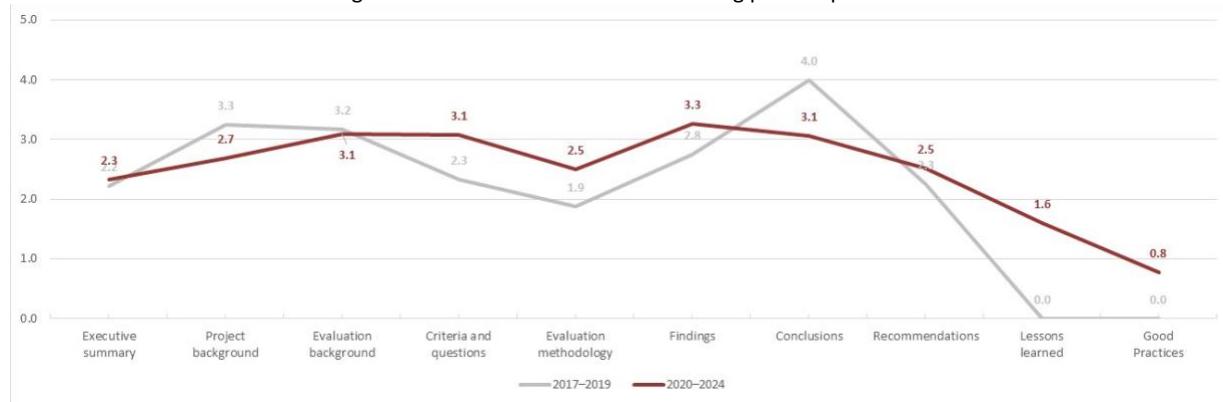
“completely correlated”, and 36 per cent of internal evaluators shared this view.<sup>23</sup> In contrast, external evaluators were more sceptical, with the majority (64%) believing the two aspects are “somewhat correlated”, and 12 per cent seeing no correlation. Another section of the questionnaire, related to the adequacy of the budget, also reflects differing views in the same direction. Evaluation managers expressed the highest level of satisfaction with the resources available for conducting evaluations, while both internal evaluators and external evaluators were more divided (for internal evaluators, 45% agreed and 55% disagreed with the statement that the budget was sufficient; for external evaluators, 42% agreed and 58% disagreed with that statement).

Some of the key informants interviewed, both evaluation managers and external evaluators, also highlighted that budgets are often inadequate, leading to compromised analytical depth and report quality. Although institutional policies recommend allocating 2–4 per cent of project budgets for evaluation, this is not consistently applied due to competing financial pressures and a lack of enforcement mechanisms. These constraints frequently result in reduced scope and limited data collection and analysis. The lack of dedicated funding also affects the selection of evaluators, sometimes prioritizing cost over expertise.

### 3.7. Analysis by component

The analysis of the average ratings for each component of the evaluation reports offers comparative insights. With an overall average rating of 2.8 points, this score falls within the upper range of “approaches requirements” and is only 0.2 points below the “meets requirements” level.

Figure 17. Arithmetic means overall rating per component



The highest-scoring component is “findings”, at 3.3, while “evaluation background”, “criteria and questions” and “conclusions” also score above 3; all are in the “meets requirements” level. The “executive summary”, “project background”, “evaluation methodology” and “recommendations” components are below the average rating but still within the “approaches requirements” level. Only “lessons learned” and “good practices” do not meet the minimum requirements, remaining at “misses requirements”.

Compared to the 2017–2019 meta-evaluation, the scores of the project background and conclusions components have notably decreased, by 0.6 and 0.9 points, respectively. In

<sup>23</sup> Internal evaluators are allocated a designated number of days by the evaluation manager to complete the evaluation, in coordination with the evaluator’s supervisor.

contrast, the scores for criteria and questions, evaluation methodology and findings have improved significantly, with criteria and questions experiencing the largest increase of 0.8 points.

Annex VIII provides a detailed analysis of the evolution of global component scores across different geographical areas.

### 1 Component 1. Executive summary

Executive summaries of evaluations were assessed to determine how well they presented seven aspects, the “items” of this analysis; the assessment of them made up 5 per cent of the global evaluation report rating. Table 10 illustrates these items and the “elements”, or content specifications on which the assessment is based; increasingly darker blue shading indicates that a higher proportion of reports achieved that particular rating for its deployment of that item.

Table 10. Overall results for component 1, executive summary

Component 1. Executive summary. Weight 5%						
Items	Elements	Rating <sup>24</sup>				
		HU	U	SU	SS	S
Explanation of the intervention background	1. Purpose 2. Objectives	14%	8%	10%	2%	60% 6%
Overview of evaluation background	1. Purpose 2. Scope 3. Intended users	11%	7%	24%	21%	33% 3%
Concise description of evaluation methodology	1. Choice of methodology 2. Data sources 3. Limitations	12%	11%	37%	19%	19% 2%
Summary of all evaluation findings	All findings included in the report	13%	4%	10%	16%	52% 4%
Summary of all conclusions	All conclusions included in the report	51%	4%	4%	9%	31% 0%
Summary of all lessons learned and good practices	All lessons learned and good practices included in the report	72%	2%	11%	4%	9% 1%
Summary of recommendations	All recommendations included	20%	4%	3%	11%	60% 1%

Note: The scores are highly unsatisfactory (HU), unsatisfactory (U), somewhat unsatisfactory (SU), somewhat satisfactory (SS), satisfactory (S), highly satisfactory (HS). A more detailed description of each rubric can be found in Table 6 (overall scoring systems).

The analysis highlights that 68 per cent of reports received a positive rating for their “explanation of the intervention background”, while “summary of recommendations” and “summary of all evaluation findings” were both rated positively in 72 per cent of reports. All three items fall within the positive spectrum. This means that, although concise, the reports included a comprehensive project description, including objectives, implementation periods and other necessary information. Likewise, the recommendations have been adequately summarized, and all findings have been included, normally organized by criteria.

In contrast, the items for which most reports received negative ratings were the “summary of all lessons learned and good practices” (85% of reports), and the “summary of all conclusions” (59%). The notably high percentage of reports that do not directly include a summary of lessons learned and good practices is striking. This often occurs because these sections are

<sup>24</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

missing from the body of the report. However, even when these sections are present in the report as a whole, lessons learned and good practices are frequently omitted from the executive summary, particularly in the case of good practices. As for the conclusions, they are often mixed in with the findings or omitted entirely from the executive summary.

The analysis identified significant variability in the scores of two evaluation items: “overview of evaluation background” and “concise description of evaluation methodology.” For the overview of evaluation background, 58 per cent of reports received positive ratings; for the description of the methodology, 60 per cent of evaluations were rated negatively. These results primarily reflect the absence or inadequate inclusion of data sources and evaluation limitations. However, compared to the 2017–2019 meta-evaluation, the quality of evaluations with regard to both traits has improved; the proportion of reports receiving a positive rating in the “overview of evaluation background” category increased by 19 percentage points (from 38% to 58%), and the proportion of reports receiving a positive rating for the description of their methodology also increased, by 31 percentage points (from 9% to 40%).

## 2 Component 2. Project background

The project background component was assessed in terms of four items, and contributed 5 per cent of each evaluation report’s total rating.

Table 11. Overall results for component 2, project background

Items	Elements	Rating <sup>25</sup>					
		HU	U	SU	SS	S	HS
Brief outline of contextual factors	1. Socioeconomic context 2. Political context	8%	4%	11%	10%	44%	22%
Clear and relevant description of key stakeholders	1. Rights holders 2. Implementing agency (or agencies) 3. Duty bearers or responsibility holders	13%	10%	18%	12%	41%	6%
Description of intervention or project logic	1. Objectives and results 2. Causal logic and assumptions	7%	13%	28%	14%	27%	11%
Funding arrangements	1. Reference to funding arrangements 2. Specific contributions of the IOM	42%	8%	10%	9%	29%	2%

Analysis of this component highlighted a significant difference from the 2017–2019 meta-evaluation. While 30 per cent of the reports during those years lacked a background section on the intervention under evaluation, all reports in the current period included this section.

Most evaluations were assessed positively with regard to the first item, “brief outline of contextual factors”, with 76 per cent of evaluations rated in the positive spectrum. These reports typically provided a comprehensive description of the context in which migration processes occur and the economic, social and institutional conditions surrounding those processes. As well, 59 per cent of reports were rated positively in terms of the second item, “clear and relevant description of key stakeholders”. These reports delivered a well-defined identification of stakeholders and their roles and responsibilities. However, in 37 out of 90 reports, there was room for improvement by further exploring the various stakeholder groups, particularly rights holders.

<sup>25</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

Fewer reports were assessed positively against the fourth item, “funding arrangements”, with 54 out of 90 reports (60%) assessed as unsatisfactory in some regard. A total of 38 reports provided no information on funding sources, while others only presented the overall project cost without specifying the IOM contribution. The quality of evaluations in this regard has worsened since the previous meta-evaluation, in which only 39 per cent received a negative rating.

The third item, “description of intervention”, showed the most even distribution of ratings for this component, with 48 per cent of reports rated negatively and 52 per cent rated positively. Incorporating (for instance) the project logical framework with indicator matrices and explanations of the intervention’s theory of change improves the quality of the evaluations.

Comparing the 2017–2019 data reveals that assessments of all reports showed improvement in terms of these items, except, as already noted, funding arrangements.

### 3 Component 3. Evaluation background

The evaluation background component was assessed in terms of three distinct items and represented 5 per cent of the total global rating of each evaluation report.

Table 12. Overall results for component 3, evaluation background

Items	Elements	Rating <sup>26</sup>					
		HU	U	SU	SS	S	HS
Explanation of purpose of evaluation	1. Purpose 2. Explanation of evaluation purpose	2%	4%	12%	14%	60%	7%
Description of evaluation scope	1. Geographic coverage 2. Time frame 3. Thematic coverage	2%	0%	13%	12%	66%	7%
List of evaluation clients and main audiences of the report	1. Intended users (such as donors or partners) 2. Intended use per client	29%	6%	14%	20%	26%	6%

Most evaluation reports – 76 out of 90 (85%) – were rated as satisfactory in terms of their description of evaluation scope. This was the most positively assessed item of this component. These reports stood out for their clear evaluation scope and well-defined objectives. Most reports were also rated positively in terms of their explanation of the purpose of evaluation, with 81 per cent of reports excelling due to their well-structured timeline, clearly articulated topic focus and comprehensive geographical coverage. These scores are similar to that received by evaluations in the 2017–2019 assessment: roughly the same proportion of evaluations assessed in this meta-evaluation received a positive rating with regard to their description of evaluation scope as did the evaluations assessed in the 2017–2019 meta-evaluation, though the proportion of evaluations receiving a positive rating regarding their explanation of purpose declined by 13 per cent.

Regarding the third item, “list of evaluation clients and main audiences of the report”, roughly the same number of reports were assessed as satisfactory as were assessed as unsatisfactory.

<sup>26</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

The percentage of reports rated in the positive spectrum with regard to this item, 51 per cent, was the same as in the 2017–2019 assessment.

#### 4 Component 4. Evaluation criteria and questions

The section of each evaluation report addressing evaluation criteria and questions was analysed with regard to three items, each of which was assessed against two elements. Assessment of this component contributed to 7.5 per cent of each evaluation report's total global rating.

Table 13. Overall results for component 4, evaluation criteria and questions

Component 4. Evaluation criteria and questions. Weight 7.5%						
Items	Elements	Rating				
		HU	U	SU	SS	S
Declaration of the evaluation criteria and justification for their use or lack thereof	1. Evaluation criteria 2. Justification of their use	1%	0%	8%	14%	49% 28%
Relevance of evaluation questions	1. Questions address the goals and purpose of the evaluation 2. Questions include cross-cutting issues	10%	0%	19%	38%	22% 11%
Inclusion of an evaluation matrix	1. Evaluation matrix, including evaluation indicators 2. Methods and sources of data per question included in evaluation matrix	31%	9%	4%	6%	49% 1%

Most evaluation reports were assessed as satisfactory in terms of the item “declaration of the evaluation criteria”, with 82 reports out of 90 (91%) rated in the positive spectrum, 7 per cent more than in the 2017–2019 assessment. Mentioning OECD-DAC criteria is considered sufficient to meet the requirements for this element. Their application is standardized among internal and external evaluators and within the Organization when drafting ToR. A significant change was the revision of the OECD-DAC criteria, which were updated in December 2019 to introduce the coherence criterion. This modification required an adaptation period until full incorporation. This change was reflected in the *IOM Monitoring and Evaluation Guidelines* in 2021, and by that year, its implementation had become widespread.

There was more variation across the evaluation reports that were assessed in terms of how well they demonstrated the relevance of evaluation questions. Though 71 per cent of reports were evaluated positively, with “somewhat satisfactory” the most common (received by 38 per cent of the reports), a key issue was that nine reports lacked evaluation questions entirely (in four cases, the ToRs were not accessible), representing a significant limitation. The weakest aspect of many evaluation reports in this regard was the limited integration of cross-cutting themes. The meta-evaluation looked for the inclusion of gender equality, RBA in programming, disability inclusion and environmental sustainability. Only one report from 2023 provided an analysis of all four cross-cutting themes.<sup>27</sup> Gender mainstreaming and the use of RBA in programming are already widely implemented (in 76 and 57 reports, respectively). Disability inclusion appeared in the evaluation questions of only nine reports, while environmental sustainability was addressed in only seven. A positive aspect is the increasing incorporation of cross-cutting themes in recent reports (2022–2024), reflecting an enhanced focus on these priorities.

<sup>27</sup> IOM Central Evaluation Division, “Thematic evaluation of IOM’s labour migration and mobility strategy and initiatives” (April 2023).

Evaluations varied even more widely in how well they included an evaluation matrix. While 56 per cent of reports were rated in the positive spectrum, 44 per cent fell into the negative spectrum. The primary reason for this was that 28 out of 90 evaluation reports lacked an evaluation matrix, and in another 8, its existence was mentioned but was not accessible in the report's annexes. An area for improvement would be to include benchmarks (reference indicators) in the matrices to facilitate well-founded value judgements. The absence of this information prevented reports from being able to achieve higher ratings. However, compared to the 2017–2019 assessment, there has been an improvement, as the proportion of reports rated at the extreme end of the negative spectrum ("highly unsatisfactory") has decreased from 58 per cent to 31 per cent, and the proportion of reports that received any kind of negative rating also decreased, from 64 per cent to 44 per cent.

## 5 Component 5. Evaluation methodology

Evaluations were analysed against eight items to assess their evaluation methods. Assessment of this component made up 7.5 per cent of the overall global rating of each evaluation.

Table 14. Overall results for component 5, evaluation methodology

Items	Elements	Rating <sup>28</sup>					
		HU	U	SU	SS	S	HS
Statement of the evaluation approach	1. Approach used 2. Justification	42%	11%	16%	8%	14%	9%
Type of analysis considered	1. Type of analysis 2. Reliability assurance and levels of evidence	2%	11%	20%	21%	34%	11%
Description of data sources	1. List of data collection methods and explanation 2. Description of different data sources	0%	1%	14%	6%	51%	28%
Sampling procedures	1. Sampling procedures 2. Justification for the selection	19%	9%	13%	24%	23%	11%
Inclusion of cross-cutting issues	1. Gender mainstreaming 2. Human rights-based approach 3. Disability inclusion 4. Environmental sustainability	23%	10%	32%	22%	9%	3%
Stakeholder participation	1. Rationale for stakeholder participation 2. Level of participation	4%	29%	26%	12%	24%	4%
Limitations of the evaluation	1. Limitations 2. Mitigation strategies	12%	2%	14%	19%	42%	10%
Description of evaluation norms and standards	1. IOM and UNEG norms and standards 2. Ethical safeguards and data protection principles	64%	1%	6%	2%	17%	10%

The highest-rated items were "description of data sources", with 76 out of 90 reports (84%) rated positively, and "limitations of the evaluation", with 64 out of 90 reports (71%) rated in the positive spectrum. Assessments of both metrics have improved compared to the 2017–2019 assessment, where 83 per cent and 56 per cent, respectively, received positive ratings. These positive ratings were largely due to the presence of evaluation matrices that effectively incorporate data collection methods and clearly outline the sources of information. Regarding the limitations of the evaluation, there was an increase in the proportion of evaluations that

<sup>28</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

included limitations and mitigation strategies (from 56% to 71%), particularly in reports from late 2020 and 2021, where higher-rated evaluations included mitigation measures to address restrictions caused by the COVID-19 pandemic.

More evaluations received negative assessments in “description of evaluation norms and standards,” with 71 per cent of reports rated as unsatisfactory, and “statement of the evaluation approach”, with 69 per cent of reports unsatisfactory. Compared to the results of the 2017–2019 assessment, the proportion of reports that were given a positive rating for their description of evaluation norms and standards increased by 12 percentage points (from 17% to 29%), while the statement of the evaluation approach was assessed positively in 31% of reports (compared to 27%), an increase of 4 percentage points. In this assessment, 58 out of 90 reports did not mention IOM and UNEG norms and standards. It is also important to note that, beyond being a protocol requirement, reports that explicitly address data protection and data handling receive particularly positive recognition. Unfortunately, in some cases, sensitive information, such as informants’ phone numbers, was included in the evaluation reports.

The low rating that many reports received in terms of the first item, “statement of the evaluation approach”, indicates a significant methodological weakness. There was often confusion between the evaluation approach – which refers to various ways an evaluation can be designed (for example, a utilization-focused evaluation, programme theory, outcome mapping) – and the qualitative and quantitative techniques used to collect information addressing evaluation indicators. As a result, 48 reports either lacked an evaluation approach or presented an incorrect one, while only 21 clearly demonstrated an understanding and justification of their chosen approach.

Assessments of the remaining four items – “type of analysis considered”, “sampling procedures”, “inclusion of cross-cutting issues” and “stakeholder participation” – were more variable. In general, the type of analysis considered and the description of data sources were rated positively for most evaluations, at 67 per cent and 84 per cent, respectively. However, with regard to the type of analysis considered, reports continued to omit explaining how triangulation processes are conducted to ensure reliability. These figures remain roughly unchanged compared to the 2017–2019 assessment. Most reports fell within the negative spectrum in terms of how well they included cross-cutting issues and how well they addressed stakeholder participation. Specifically, in terms of their inclusion of cross-cutting issues, 66 per cent of the reports received negative ratings, with 30 reports not mentioning cross-cutting issues in their methodology or referring only to gender equality. Despite this, there is relative improvement compared to the 2017–2019 assessment, where only 4 per cent of reports were assessed positively in this regard. This suggests that while progress has been made in integrating cross-cutting issues, significant room for improvement remains.

The quality of sampling procedures has declined compared to the reports assessed in the 2017–2019 assessment. In the current period, 59 per cent of reports received positive ratings for their sampling procedures, down from 69 per cent in the previous meta-evaluation. Sixteen reports did not include a sampling section, and only 10 achieved the highest scores. Reports often described informants without explaining how gender balance or representativeness in the project was ensured, resulting in inadequate justification of selection processes.

Evaluation reports have in general improved since the 2017–2019 assessment in terms of how well they address stakeholder participation, with 41 per cent of reports receiving a positive assessment (an increase of 12 percentage points, from 29%). However, in the remaining 59 per cent of reports that did not receive positive ratings, stakeholder participation analysis remained superficial. In most cases, there was no in-depth assessment of stakeholder involvement in the evaluation process beyond a basic description of their role in the intervention.

These issues were also raised by evaluation managers and evaluators in the interviews. A significant number of them underlined that, in some cases, evaluation questions and methodologies were not always well-designed, directly impacting the assessment's accuracy and usefulness. It is common to find questions that are overly generic or copied from previous evaluations without proper adaptation to the project's context. This was reinforced by the automatic application of OECD-DAC criteria without a strategic reflection on their relevance. Other methodological weaknesses, such as unclear sampling strategies or insufficient data triangulation, further compromised the credibility of findings, resulting in reports that lack analytical depth. Several key informants stated that when methodologies are carefully designed and adapted to the specific evaluation context, reports tend to be more insightful and produce practical and actionable recommendations.

## 6 Component 6. Evaluation findings

Four items were analysed to assess the quality of the evaluation findings of each evaluation report. Of all the components, this one had the highest relative weight (25%) in determining the overall rating of each evaluation.

Table 15. Overall results for component 6, evaluation findings

Component 6. Evaluation findings. Weight 25%						
Items	Elements	Rating <sup>29</sup>				
		HU	U	SU	SS	S
Completeness	1. All evaluation criteria and questions are addressed	0%	3%	1%	20%	51%
	2. Findings aligned with purpose, questions and approach					
Robustness	1. Findings justified by multiple lines and levels of evidence	0%	1%	17%	31%	34%
	2. Disaggregation of data by key variables, indicators and benchmarks as per the initiative's M&E framework					
Identification of causal factors	1. The causal factors explaining the achievement of results or lack thereof are identified and justified	1%	1%	24%	21%	40%
	2. Description of unintended effects					
Cross-cutting issues addressed	1. Gender mainstreaming	7%	13%	38%	33%	6%
	2. Human rights-based approach					
	3. Disability inclusion					
	4. Environmental sustainability					

The item that was most often addressed in a satisfactory manner across the evaluation findings of the evaluation reports was “completeness”, with 95 per cent of reports rated in the positive spectrum. This indicates that all evaluation questions and criteria had been adequately

<sup>29</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

answered and were aligned with the evaluation goals. Compared to the reports assessed in the 2017–2019 meta-evaluation, there has been an improvement of 12 percentage points, from 84% to 96%. A qualitative analysis of reports placed in the negative spectrum identified several limitations: challenges in evaluating efficiency due to insufficient budget details, difficulties tracing responses to specific questions due to a lack of alignment between answers and queries, and unanswered questions, particularly regarding the intervention’s design.

Most evaluations were also assessed positively in terms of their robustness and their identification of causal factors, with 82 per cent of reports receiving a positive rating in terms of their robustness and 73 per cent of reports receiving a positive rating regarding their identification of causal factors. This makes “identification of causal factors” the item of this component that showed the greatest improvement, an increase of 25 percentage points (from 48% to 73%), since the 2017–2019 assessment. On the other hand, there was only a 4 percentage point increase (from 78% to 82%) in the proportion of reports receiving a positive assessment of the robustness of their evaluation findings, as compared to reports assessed in the 2017–2019 meta-evaluation. Regarding robustness, high-scoring reports stood out due to their data disaggregation by indicator, with the best including numerical data (figures) and graphical representations, drawing from multiple data sources. Regarding the identification of causal factors, the highest-rated reports thoroughly analysed causal relationships between key components and assessed potential positive or negative impacts. However, there is still room for improvement in constructing causal explanations linking actions with results. The weakest aspect of this item was the treatment of unforeseen results, which were not included in the reports.

The item that most reports struggled to address was “cross-cutting issues addressed”, with 58 per cent of reports analysed receiving a negative assessment. This weak performance stems from the inadequate integration of cross-cutting issues into both evaluation criteria and methodology. Without embedding these issues into evaluation questions – either as a dedicated criterion or within the OECD-DAC framework or other applied methodologies – obtaining robust and meaningful findings becomes challenging. Furthermore, when evaluation methodologies lack techniques to capture cross-cutting issues, extracting relevant data is nearly impossible. As a result, most reports provided only fragmented findings on gender equality and the use of RBA in programming, while other cross-cutting themes – such as disability inclusion and environmental sustainability – remained insufficiently assessed. This lack of focus was already evident in the ToRs: among the 58 analysed, only 8 mentioned disability inclusion, and just 3 addressed environmental sustainability. However, despite these shortcomings, the total percentage of evaluations with a satisfactory rating improved from 24% to 42%, an 18 percentage point improvement from the 2017–2019 meta-evaluation to the current one.

## 7 Component 7. Conclusions

The meta-evaluation scoring tool evaluated two items to analyse the quality of the conclusions of each evaluation report. This component accounted for 20 per cent of the total global rating given to each evaluation.

Table 16. Overall results for component 7, conclusions.

Component 7. Conclusions. Weight 20%							
Items	Elements	Rating					
		HU	U	SU	SS	S	HS
Value	1. Conclusions are clearly linked to findings 2. Provide insights and add value to related findings	4%	10%	13%	27%	40%	6%
Reasoned	1. Conclusions reflect reasonable critical thinking and evaluative judgements 2. Judgements are to the extent possible objective	4%	9%	17%	23%	38%	9%

The first item, “value”, examined the relationship between conclusions and findings, and looked at the added value that conclusions provide. Analysis found that 72 per cent of reports were assessed positively (65 reports out of 90 were rated “satisfactory”, leaving 25 reports in the negative spectrum), a percentage very similar to that observed for the same item in the 2017–2019 meta-evaluation.

The best-rated reports were characterized by conclusions clearly linked to findings, following the evaluation criteria or questions. These conclusions were also detailed, concise and connected to the most relevant knowledge from the analysis. Reports with room for improvement sometimes omitted the conclusions section, jumping directly from findings to recommendations. In other cases, the reports needed to establish a clearer link between findings and conclusions, moving beyond merely repeating or narrating findings by adding value through connections to other information. It is also important that conclusions reflect findings that pose limitations for the project, such as the lack of a gender perspective or the failure to mainstream other cross-cutting issues.

Assessment of the “reasoned” item evaluated whether conclusions were grounded in sound critical thinking and maintained objectivity as much as possible. Seventy per cent of reports received a positive rating, 10 percentage points less than in the 2017–2019 meta-evaluation. This leaves 30 per cent of reports in the negative spectrum.

The best-rated reports presented detailed conclusions supported by critical analysis derived from earlier analysis and triangulation, making them considerably objective and well balanced. These conclusions were typically well written, reflected sound and well-structured value judgements, and demonstrated rigorous reasoning.

## 8 Component 8. Recommendations

The recommendations component was analysed in terms of four items, and accounted for 20 per cent of the total weight of each evaluation’s global rating.

Table 17. Overall results for component 8, recommendations

Component 8. Recommendations. Weight 20%							
Items	Elements	Rating <sup>30</sup>					
		HU	U	SU	SS	S	HS
Clarity	1. Are clear and concise						
	2. Are based on findings and conclusions of the report	2%	6%	8%	22%	51%	11%
Relevance	1. Address key issues and are useful, tied to the object and purposes of the evaluation						
	2. Address cross-cutting issues	4%	4%	33%	36%	18%	4%
Responsibility	1. Specify who is called upon to act						
	2. Identify means for achievement	18%	14%	44%	14%	8%	1%
Actionability	3. Specify priority or importance						
	1. Indicate the specific courses of action needed to remedy or continue with the current situation	26%	10%	21%	27%	12%	4%
	2. Provide a time frame						

In terms of clarity, relevance and actionability of the recommendations, the results from the previous meta-evaluation appear more satisfactory. However, there was an improvement in terms of responsibility.

The item for which the most reports received a satisfactory rating was “clarity”, with 84 per cent of evaluations rated in the positive spectrum and 16 per cent in the negative. This indicates that clear and concise recommendations were aligned with findings and conclusions. However, this marks a drop of 10 percentage points from the ratings achieved in the 2017–2019 meta-evaluation, signalling that while clarity persists, some recommendations now tend to be overly vague, lack specificity, or fail to connect directly with findings and conclusions.

Fewer reports received a satisfactory rating for the relevance of their recommendations, with 58 per cent of reports receiving a positive rating and 42 per cent a negative. This represents an improvement compared to the 2017–2019 report, where only 44 per cent of reports were in the positive range, and 56 per cent were negative. The primary reason for the many reports that rated poorly in terms of the relevance of their recommendations was their failure to address cross-cutting issues. At best, these issues were briefly mentioned, and often restricted to gender-related aspects.

The areas of the recommendations where most reports have the greatest potential for improvement, as reflected by their lower scores, are the traits “responsibility” and “actionability”. Concerning the responsibility of recommendations, 77 per cent of reports were classified within the negative spectrum. However, this is a significant improvement since the 2017–2019 meta-evaluation, in which 93 per cent of reports were categorized within the negative spectrum. The primary reason for this low score is that 68 out of 90 reports did not identify the necessary means to implement recommendations or define their priority level. Regarding the actionability of their recommendations, 51 out of 90 reports (57 per cent) fell into the negative spectrum, often because they failed to specify recommendation time frames or the necessary actions for their implementation. Compared to the 2017–2019 assessment, there is once again an improvement, in this case of 27 percentage points.

## 9 Component 9. Lessons learned

<sup>30</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

The component addressing lessons learned was analysed in terms of three items, each of which was assessed against two elements. This component was worth only 2.5 per cent of each evaluation report's total rating.

Table 18. Overall results for component 9, lessons learned

Component 9. Lessons learned. Weight 2.5%						
Items	Elements	Rating <sup>31</sup>				
		HU	U	SU	SS	S
Relevance	1. Lessons learned are derived from findings or the evaluation process	42%	3%	14%	20%	19%
	2. Lessons learned represent a relevant or new piece of information to be considered in the future					
Delimitation	1. The lessons capture the context from which they were derived	42%	9%	18%	24%	6%
	2. Lessons learned target specific users					
Applicability	1. The lessons specifically suggest what should be avoided in future contexts to guide action	46%	7%	13%	13%	20%
	2. The lessons include causal factors					

The assessment of each of the three items for this component showed that the evaluation reports had significant room for improvement. It is important to note that 37 out of 90 reports did not include this section, despite it being clearly specified in most ToRs and tools shared by IOM. This clearly impacted the percentage of reports receiving a negative rating.

Regarding the first item “relevance”, 60 per cent of reports received negative ratings. However, this marked an improvement from 75 per cent in the 2017–2019 assessment. The analysis revealed conceptual confusion between lessons learned and good practices, as well as between lessons learned and recommendations. Lessons learned were frequently conflated with recommendations, treated as secondary or repetitive extensions rather than distinct insights derived from programme experiences. Sometimes, lessons learned did not provide value beyond the information already shared.

The analysis of the evaluation reports in terms of “delimitation” reveals that 69 per cent of reports were rated negatively. As with the relevance of evaluations’ lessons learned, this was a noteworthy improvement of 16 percentage points over the evaluations assessed in the 2017–2019 meta-evaluation, in which 85 per cent of reports were rated negatively. The primary issue was the lack of contextualization for lessons learned and the failure to specify their intended audience.

In terms of the third item, “applicability”, 66 per cent of evaluations received ratings in the negative spectrum, a notable improvement (24 percentage points) from 89 per cent in the 2017–2019 meta-evaluation. Most lessons learned did not include causal factors to explain them, nor did they provide guidance on how to apply them. Only one out of 90 reports achieved a highly satisfactory rating,<sup>32</sup> making it a rare exception.

## 10 Component 10. Good practices

<sup>31</sup> The percentages shown in this Table may not sum 100% due to rounding of decimal places.

<sup>32</sup> Specifically, the evaluation report “[Evaluación final de la estrategia de fortalecimiento institucional para el despliegue y funcionamiento territorial articulado del sistema integral de verdad, justicia, reparación y no repetición](#)” (2021).

The component addressing good practices was analysed against three items, each of which was assessed in terms of two elements, with a relative weight of only 2.5 per cent of each evaluation report's total rating.

Table 19. Overall results for component 10, good practices

Component 10. Good practices. Weight 2.5%						
Items	Elements	Rating <sup>33</sup>				
		HU	U	SU	SS	S
Delimitation	1. The good practices concisely capture the contexts from which they were derived 2. The good practices specify target users	70%	1%	8%	11%	10% 0%
Applicability and replicability	1. The statements describe how the good practices should be implemented and by whom 2. The good practices explore the applicability in different contexts	76%	3%	8%	7%	7% 0%
Impact	1. The good practice demonstrates a link to specific impacts 2. The different impacts identified are viable and possible (realistic)	73%	2%	10%	6%	9% 0%

As with the analysis of lessons learned, the quality of the evaluation reports in terms of their analysis of good practices showed significant room for improvement. It is important to highlight that 61 out of 90 reports did not include this section. This absence contributes to the percentage of reports that received ratings within the negative spectrum.

Seventy-nine per cent of reports received a negative rating for how they addressed the item “delimitation”, with 70 per cent of reports rated as highly unsatisfactory. This percentage was similar to the 2017–2019 assessment (80%), indicating no improvement in this regard. As noted in the analysis of the previous component, good practices were often confused with recommendations or with lessons learned, adding little value beyond the ideas already presented in the reports’ findings or conclusions. In very few cases was the context described, and the target audience specified.

The second item, “applicability and replicability” showed the most significant need for improvement, with 87 per cent of reports falling within the negative spectrum, mirroring the percentage from the previous meta-evaluation.

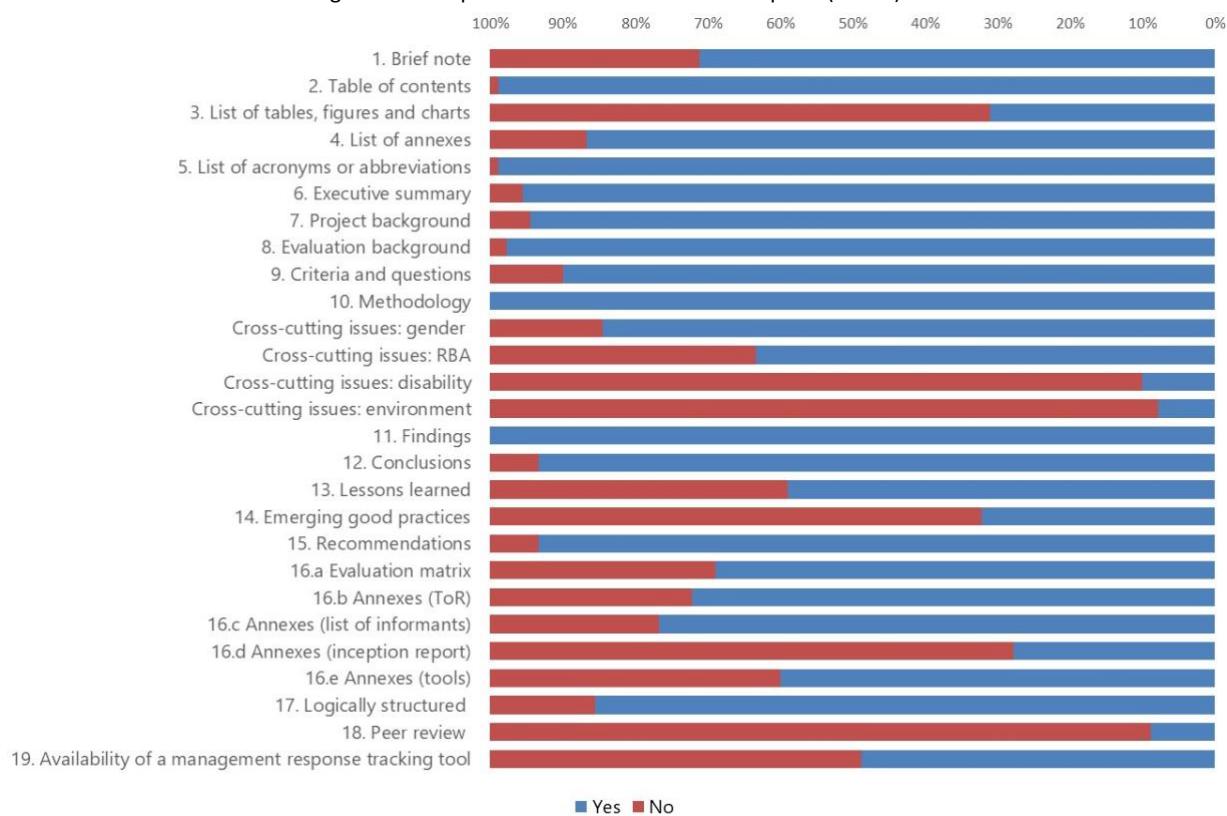
Finally, 86 per cent of reports were rated in the negative spectrum for the “impact” of their good practices, a deterioration of 8 percentage points from the 2017–2019 report. Very few reports included this section, and were thus unable to identify positive impacts.

### 3.8. Comprehensiveness of the evaluation reports

This section turns from the first dimension of analysis to the second, an analysis of the comprehensiveness of the evaluation reports. This assessment was based on the presence or absence (using a dichotomous yes/no scale) of 19 elements put forward in the scoring tool.

<sup>33</sup> The percentages shown in this Tables 10 to 19 may not sum 100% due to rounding of decimal places.

Figure 18. Comprehensiveness of evaluation reports (n = 90)



Six of the evaluated elements pertain to the report's presentation. Out of the 90 reports reviewed, nearly all included a table of contents and a list of acronyms. Additionally, 78 reports featured a list of annexes and a logical structure, while approximately 64 reports included a brief note. In contrast, only 28 reports provided a list of tables and figures.

Regarding the annexes included in the report, 69 reports included a list of informants, 65 attached the ToR, and 54 provided the tools developed for the research. However, only 25 reports included an inception report, a fundamental element for complementing and deepening key methodology and report composition aspects, such as the evaluation matrix.

The remaining 13 elements are linked to the content that an evaluation should include. Eighty-one reports contained an executive summary, project background, criteria and questions, methodology, findings, conclusions and recommendations. The coverage of cross-cutting issues, both in the evaluation questions and methodology, showed significant differences, which influenced the results in the component-based analysis. While the cross-cutting issue of gender equality was present in 76 of the reports, RBA featured in only in 57 reports, disability inclusion was included in just 9 reports, and environmental sustainability appeared in only 7 reports. Lastly, following the same downward trend, lessons learned were included in 53 reports, but emerging good practices in only 29 reports. Finally, 50 per cent of the reports featured a management response tracking tool, but only 8 explicitly mentioned that the report was peer reviewed.

The comparison with the 2017–2019 assessment of this dimension shows improvements such as the presence of an evaluation brief, tables, figures, charts, acronyms or abbreviations,

executive summary, criteria and questions, methodology and lessons learned. The rest have remained relatively unchanged.

IOM staff and external consultants emphasized that the quality assurance tools played a significant role in methodological harmonization, aligning and standardizing content between reports. Survey results indicated that 59 per cent of IOM staff find IOM guidelines and materials moderately sufficient and clear, though only 12 per cent appear completely satisfied with them. External evaluators, on the other hand, hold a more mixed view: 44 per cent consider them very useful, 38 per cent find them somewhat helpful, while 14 per cent have not reviewed them at all.

Interviews revealed differing perspectives on the usefulness of IOM methodological guidelines and quality assurance tools. Evaluation managers often regard these materials as useful frameworks, but in some cases they struggle with their practical application due to complexity, length and inconsistent adoption. On one hand, evaluators – especially external ones – may perceive them as rigid or overly standardized, sometimes applying them mechanically without fully adapting them to the evaluation context and the specific requirements outlined in the ToR. On the other hand, communication and coordination between evaluators and IOM staff vary significantly. While some evaluators benefited from strong engagement with regional offices and PMs, others operated with limited guidance and fragmented access to standardized templates, quality assurance tools and other relevant background and data sources. When coordination is proactive, findings tend to be more nuanced and better contextualized.

All these elements may have influenced the alignment and comprehensiveness of the evaluation reports.

### 3.9. Comprehensiveness of the terms of reference

The comprehensiveness of the ToRs accompanying the evaluation reports was analysed based on 14 sections or elements from the previous meta-evaluation 2017–2018,<sup>34</sup> with the section on evaluation questions further divided into four cross-cutting issues. This assessment was conducted on all the 58 ToRs attached to the 90 evaluation reports reviewed. Once again, a dichotomous yes/no scale was used to indicate the presence or absence of each element.

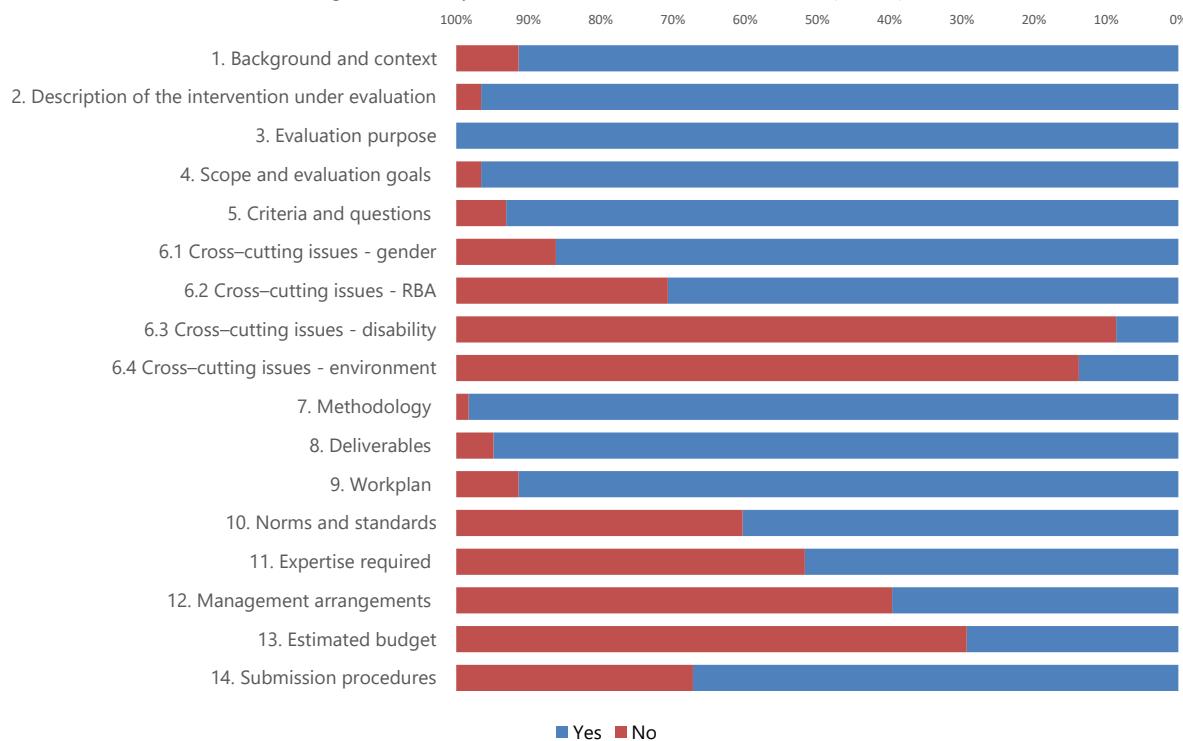
The results by ToR section (see Figure 19) indicate that the sections with the least presence (below 50%) in the ToRs were: “cross-cutting issues – disability” (5 ToRs), “cross-cutting issues – environment” (8 ToRs), “estimated budget” (17 ToRs), and “management arrangements” (23 ToRs). These four sections are included in the “Guidance on quality management of IOM evaluations” (2022), although the specific cross-cutting issues of environmental sustainability and disability inclusion are not detailed but fall under the general category of “other cross-cutting issues”. This lack of definition may contribute to their omission in ToRs.

---

<sup>34</sup> See the dimension 2 tool included in Annex 3. The tool was adapted by Artival Research & Evaluation based on the IOM checklist is of evaluation of terms of refence, referenced in the [IOM Monitoring and Evaluation Guidelines](#) (Geneva, IOM, 2021) for the “[Meta-evaluation of IOM’s internal and external evaluations \(2017–2019\)](#)” (April 2020).

The next group of sections, found in 50–69 per cent of the ToRs, includes “norms and standards”, “expertise required” and “submission procedures”, all well-represented in the checklist. Between 70 and 90 per cent of the ToRs included the cross-cutting issues of gender equality and RBA in programming, which were clearly detailed in the checklist. The remaining sections (“background, context”; “description of the intervention under evaluation”; “evaluation purpose”; “scope and evaluation goals”; “criteria and questions”; “methodology”; “deliverables”; and “workplan”) appeared in nearly 100 per cent of the ToRs.

Figure 19. Comprehensiveness of terms of reference (n = 58)

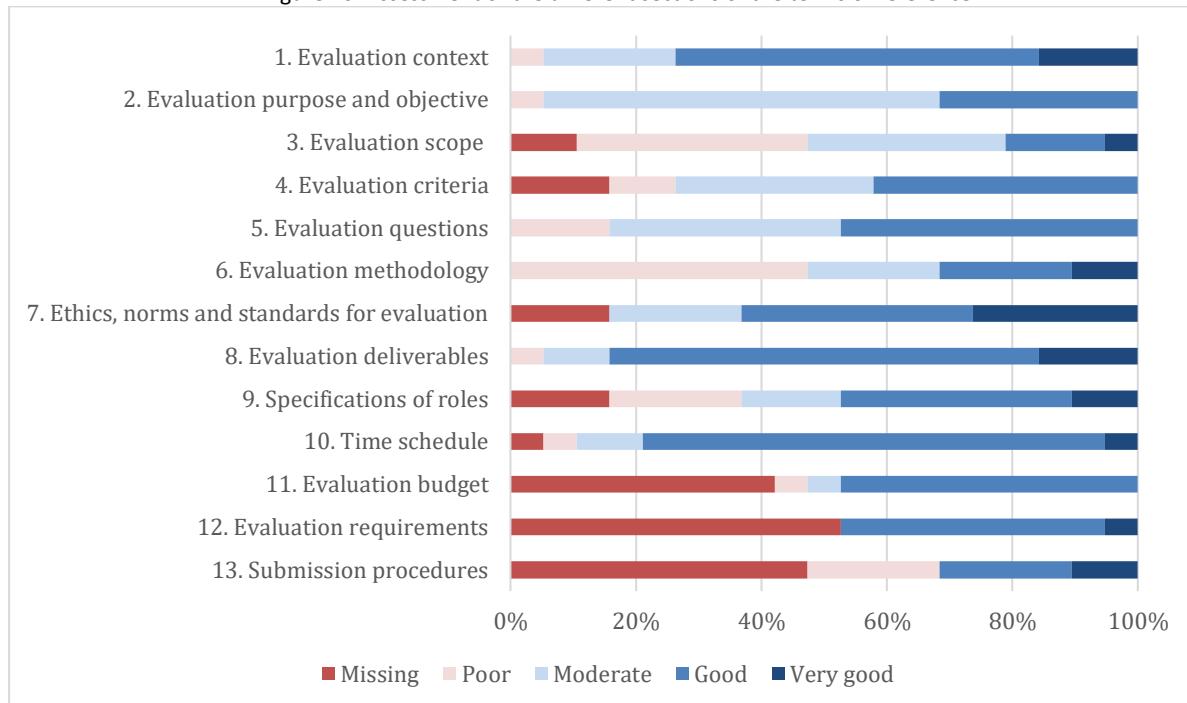


An additional, deeper analysis compared a smaller sample of 19 ToRs (see the sampling section) against the quality standards established in the revised IOM checklist for terms of reference.<sup>35</sup>

The analysis of the quality of each ToR has enabled: (1) an assessment of each section of the checklist to identify areas for improvement; (2) the acquisition of a score that, when compared to the evaluation report’s score, assists in determining any relationship between them; and (3) a better understanding of the connection between the effort put into developing guidelines and tools and the evolution of evaluation report quality, based on the yearly trends in scores following their publication.

<sup>35</sup> The revised checklist is available in the “[Guidance on quality management of IOM evaluations](#)” (September 2022). The parameters of the analysis (that is, the sections of the ToR that are being examined) differ to some extent from those included in Annex 3 used for the analysis above.

Figure 20. Assessment of the different sections of the terms of reference



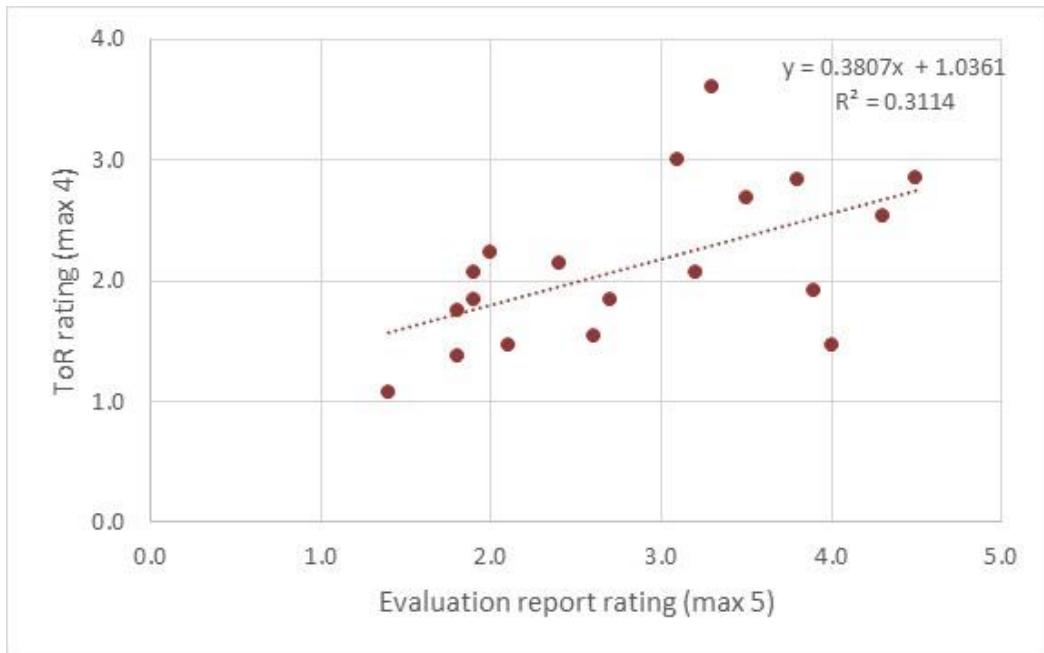
A detailed analysis of the completeness of ToRs indicates that the sections with results in the positive spectrum (moderate, good, and very good) for more than 70 per cent of the ToRs are: “evaluation context” (95%), “evaluation purpose and objective” (95%), “evaluation criteria” (74%), “evaluation questions” (84%), “ethics, norms and standards” (84%), “evaluation deliverables” (95%) and “time schedule” (89%).

Conversely, the sections with the highest proportion of results in the negative spectrum (missing and poor) are: “evaluation scope” (47%), “evaluation methodology” (47%), “specification of roles” (37%), “evaluation budget” (47%), “evaluation requirements” (53%) and “submission procedures” (68%).

When analysing the methodology section, it is advisable to require the evaluation team to define the evaluation approach clearly and to ensure that the data is collected in a manner that follows a human rights-based approach and that respects gender perspectives, and that it disaggregates according to social criteria. Sometimes, it is essential to ensure that the evaluation team can organize the evaluation questions outlined in the ToRs and limit them to a number that effectively addresses the evaluation objectives without exceeding a reasonable amount. The other sections with low scores mainly suffer from their omission in the ToRs, which should be addressed through a review based on the checklist. The remaining sections in the negative spectrum primarily reflect their omission in the ToRs, as drafting them typically does not pose difficulties.

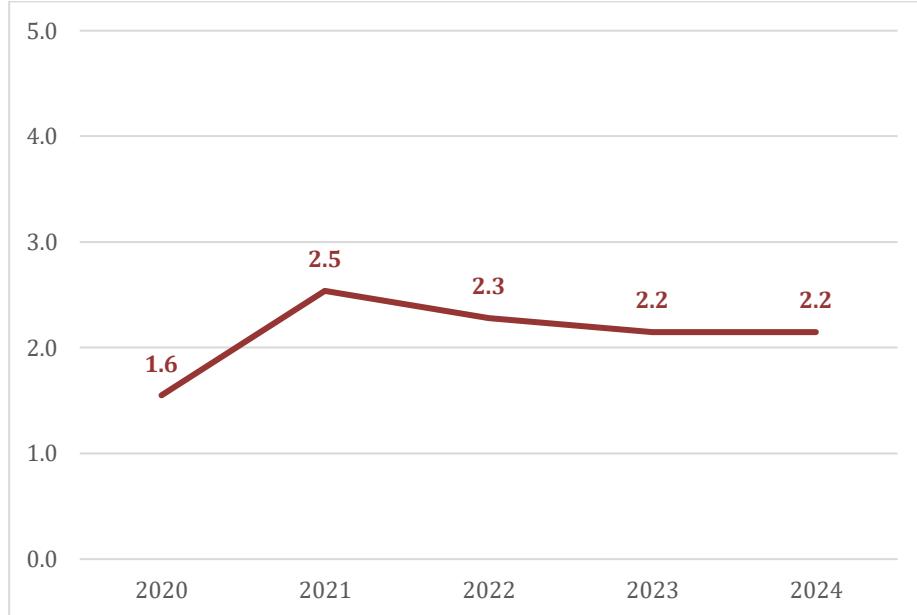
Figure 21 plots the relationship between the quality of evaluation reports and the quality of their ToRs.

Figure 21. Relationship between global scoring terms of reference and global scoring reports



The global scoring of the reports (maximum 5) is located on the x-axis, and on the y-axis, the global scoring of the ToRs (maximum 4) is displayed. A positive correlation is observed between the scores of ToRs and the scores of evaluation reports. However, the dispersion of points suggests that this relationship is not perfectly linear. The slope of the trend line is not very steep, indicating a certain degree of positive correlation: the quality of the ToRs does hold some significance, although is not the only factor affecting the quality of the reports.<sup>36</sup>

Figure 22. Global average quality score of the terms of reference by years



Finally, the quality of the ToRs significantly improved from 2020 to 2021, coinciding with the publication of the guidelines in 2021. Although this momentum weakened somewhat in

<sup>36</sup> ( $R^2 = 0.3114$ ). The statistical test confirmed that this relationship was highly significant ( $p\text{-value} = 4.73 \times 10^{-4}$ ), meaning it is very unlikely to be due to chance. This suggests that when ToRs are good the quality of the evaluation tends to improve.

subsequent years, the average score remains moderate, above 2 out of 5 (see Figure 22). However, since these results are based on a small sample, caution should be exercised when drawing conclusions.

The survey results point in the same direction. According to the evaluation managers, evaluation ToR quality has improved over the past five years. The majority (61%) agreed with a statement to this effect, 22 per cent somewhat agreed, and a smaller proportion (8%) were strongly convinced. The remaining 8 per cent preferred not to provide an opinion.

This perception was widely shared by interviewees, who viewed the trend as positive but identified some ongoing challenges. They noted that introducing standardized templates and offering continual support from the regional offices have contributed to greater consistency in the structure of ToRs and in the adherence to institutional guidelines. However, the involvement of regional offices during the design and fine-tuning process remains uneven, primarily depending on whether PMs proactively seek their input. Additionally, it was mentioned that some ToRs duplicated text and generic information from prior assessments instead of being tailored to the specific context (for example, related to the OECD-DAC criteria or the evaluation questions). This lack of critical assessment and contextual adaptation diminishes the strategic focus of these contractual guidelines. Training initiatives have helped enhance technical capacities, but high staff turnover and the limited technical skills and human resources of some country offices have significantly hindered harmonization and quality standards. Furthermore, stakeholder consultation in the drafting process is often restricted, with donors, beneficiaries and implementing partners rarely involved in shaping evaluation questions and criteria.

### 3.10. Comprehensiveness of inception report

Three inception reports were analysed using the IOM quality control tool.<sup>37</sup> As outlined in the methodology section, one report was selected for analysis from each of 2020, 2022, and 2023. The content of the different inception report sections was examined.

#### Review by sections

##### *Introduction*

The 2020 report excelled by including a table explicitly detailing key components. The 2022 report included a table of contents but omitted summaries of basic information. And the 2023 report showed further regression, lacking a detailed table of contents and the summary.

##### *Evaluation context*

The analysis of inception reports from 2020 to 2023 reveals evolving but inconsistent adherence to key evaluation requirements. The 2020 inception report provided basic contextual explanation and a simple logical framework but omitted a theory of change, details on links with other interventions or institutions, and specifics about implementing entities. In contrast, the 2022 report improved by including the institutional context and a matrix outlining timing, budget, donor involvement, outcomes and outputs for each of the five projects within the intervention. However, it still lacked a theory of change. The 2023 inception report,

---

<sup>37</sup> IOM Central Evaluation Unit, “[Guidance on quality management of IOM evaluations](#)” (September 2022).

focused on a thematic evaluation, detailed the division under review using strategic documents and outlined the two main objectives being evaluated.

#### *Evaluation purpose, criteria and questions*

The 2020 inception report did not mention the evaluation scope and instead focused on explaining changes in objectives and fieldwork visits, justifying these modifications based on the programme's evaluability limitations. The evaluation matrix, included in the annexes, featured high-quality and well-structured questions that were numerous and well-linked to indicators. It aligned with the results established in the project's results framework and established a robust triangulation system between quantitative and qualitative techniques, incorporating a diverse range of research methods. The matrix included specific criteria for cross-cutting issues, covering disability, gender and RBA. Overall, the matrix was highly comprehensive and well executed.

The 2022 inception report addressed the objectives pursued by the evaluation within the main text; however, it did not explicitly state who the users of the evaluation were or how its results were expected to be used. It included a very basic evaluation matrix in the annexes, which, while covering all necessary components (except for standards), lacked depth and detail. A criterion for cross-cutting issues was introduced, but it was limited to gender and the environment. Data sources and methods were sparse, repetitive and did not demonstrate clear triangulation, further weakening analytical rigour. Notably absent was any consideration of using project monitoring data to assess objective achievement.

The 2023 inception report incorporated all the aspects that needed to be included in this section, explaining the reasons for adapting the TOR request. However, while the matrix presented well-developed evaluation questions and subquestions, it lacked indicators and applied the same three data collection techniques to all criteria. It included the four cross-cutting issues, but only within the first three criteria, omitting them from the impact and sustainability criteria. This made it the weakest matrix of those provided in the three inception reports.

#### *Methodology*

The 2020 inception report comprehensively presented a methodology . It disaggregated each technique by informant, applied multiple data collection methods and outlined how the data would be analysed. Additionally, it included a diverse sampling approach based on the techniques to be applied, ensuring gender proportionality. The section on limitations and mitigation was also included.

The 2022 inception report included all required elements but was less detailed than the 2020 example. It proposed only two techniques – document review and interviews – and the sampling was limited to a few informants, but the limitations and their mitigation were well explained.

The 2023 inception report presented an extensive methodology section, covering all required components. An interesting addition was the inclusion of a case study to complement document review, interviews and surveys.

#### *Workplan*

All three inception reports included a chapter covering the required information for this section, with the 2020 report being the most comprehensive.

#### *Annexes*

The essential annexes that must be included are the evaluation ToR, the evaluation matrix, and the data collection instruments. The latter two are found in all three inception reports, but the ToRs are only included in the 2020 report.

#### *Overall analysis*

These are preliminary insights are based on a very small sample of inception reports. To form definitive conclusions, further analysis with larger samples will be required.

This analysis suggests that inception reports contain two types of information, strategic and operational. Strategic information includes the purpose, criteria and questions, particularly emphasizing the evaluation matrix and methodology. The operational information addresses broader organizational aspects, such as context and workplan. It is crucial for the strategic components to be clearly articulated and properly reviewed. The 2020 inception report, which featured a high-quality evaluation matrix and methodology, also achieved the highest overall score of 3.9. While a direct correlation between these variables is not claimed, this example point towards a potential link between the quality of these strategic elements and a strong evaluation report. However, further analysis with a more comprehensive sample would be necessary to confirm this relationship.

No clear connection could be established between the publication of the guidelines and the improvement in inception report quality, as a sample of three inception reports is insufficient to observe trends over time. However, all three inception reports present a high level of quality, which is reflected in the global scores, as all fall within the “meets requirements” group.

## **Findings on the evaluation process**

The results presented in this section emphasize the evaluation process, from design to reporting. The findings primarily stem from the online survey analysis and key informant interviews, which are supplemented by a desk review of essential documents related to the internal evaluation process. This section addresses meta-evaluation questions 2 and 3 (see Table 4).

### **Key findings**

12. The decentralized approach to evaluation design results in variations in scope and methodological robustness, and in a lack of consistency. While institutional policies and donor requirements provide a general framework, evaluations are often planned on an ad hoc basis, with limited mechanisms to ensure systematic planning, development and follow-up. This affects short-term or lower-budget projects, where evaluations are less frequent, and their strategic orientation is not always clearly defined.

13. The development of ToRs remains inconsistent, with significant differences in clarity, focus and methodological precision. Although PMs oversee the drafting process, the criteria

used are not always tailored to the specific context of each evaluation. Many ToRs include all OECD-DAC criteria without assessing their actual applicability, leading to broad scopes that can dilute the depth of analysis. The review process for ToRs is not systematically enforced, with the irregular involvement of regional offices contributing to these variations.

14. Quality assurance tools, such as standardized templates and review checklists, have contributed to greater structure in evaluation processes, yet their use is inconsistent across different offices. While these tools align with international standards, their accessibility and integration into workflows remain limited. In some cases, they are applied selectively; in others, they are bypassed due to competing priorities or a lack of familiarity with their content.

15. The implementation of evaluations is shaped by logistical and procedural factors, including available resources, time constraints and internal coordination. While some evaluations benefit from careful planning and oversight, others are carried out with limited methodological guidance, or late in the process. The level of involvement from regional M&E officers and technical staff varies widely, affecting coherence in execution and quality control throughout the process.

16. The structure and analytical depth of evaluations vary considerably, influenced by differences in methodological approaches and data collection strategies. While some reports demonstrate strong adherence to evaluation standards, others exhibit gaps in documentation, transparency and methodological traceability. The selection of key informants and reference documents is not always systematic, which can affect the diversity of perspectives incorporated into the analysis and increase the risk of bias.

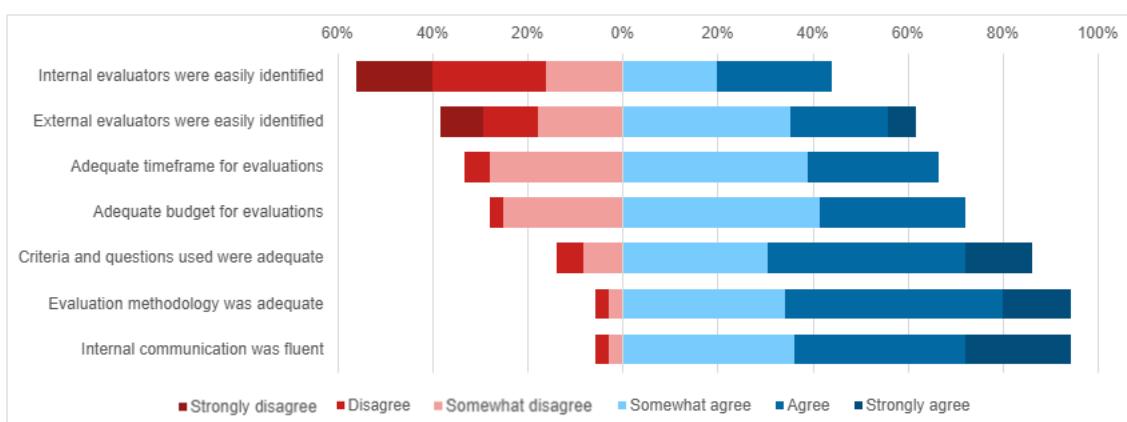
17. The evaluation function remains poorly institutionalized, lacking mandatory requirements for systematic implementation such as standardized review and approval workflows, or designated leadership. High staff turnover and the resulting expertise gaps have impacted the quality of evaluations, while restructuring processes have fragmented responsibilities and diluted evaluation efforts within broader functions. Consequently, evaluations are conducted inconsistently and with varying levels of quality and strategic alignment, limiting their potential to influence organizational priorities and programmatic decisions.

18. Recent efforts to enhance evaluation visibility and accessibility have created opportunities to strengthen the role of evaluation in organizational learning and decision-making. Initiatives like promoting the widespread use of evaluation briefs and increased engagement with donors and Member States have improved dissemination. Concurrently, discussions on structured follow-up mechanisms indicate a growing recognition of the need for systematic and supervised use of findings. Additionally, the increased emphasis on knowledge management, in particular lessons learned and good practices, and regional thematic assessments presents a path to increasing the strategic relevance of evaluation, ensuring a more effective contribution to programmatic improvements and the design of new evidence-based proposals.

Discussions with IOM staff and external evaluators revealed that, in recent years, evaluation processes have benefited from increased standardization and improved methodological guidance. The introduction of structured templates and expanded technical support have contributed to greater evaluation design and reporting consistency. While challenges persist, these developments provide a basis for strengthening the strategic value of evaluations, improving analytical standards and better integrating findings into decision-making. Additionally, growing institutional awareness of the strategic purpose of evaluations and the active inclusion of IOM staff in the data collection and analysis processes have led to the gradual adoption of more harmonized and evidence-based practices.

Figure 23 shows the responses provided by the evaluation managers to the online survey, which coincide with some of the conclusions and ideas shared by the key informants consulted.

Figure 23. Level of agreement of surveyed evaluation managers with statements about the evaluation process



Overall, a strong consensus exists on the adequacy of the evaluation methodology, criteria and questions used, as well as internal communication, with most respondents expressing positive views. This suggests that, despite some operational challenges, the technical foundation of evaluations is generally seen as adequate. However, considerable disagreement exists regarding the identification of evaluators, especially regarding the extent to which internal evaluators were easily identified, with many key informants reporting difficulties. This highlights potential gaps in recruitment processes, in the availability of qualified personnel, or in clarity regarding selection mechanisms. Similarly, budget and time constraints remain significant challenges, as many respondents disagreed that the available budget and evaluation timeline were adequate.

The analysis has identified four areas that have had a significant impact on the evaluation process and the quality of the reports:

- The increased use of quality assurance tools has led to greater consistency in evaluation formats and processes. Standardized templates, methodological guidelines and review checklists have helped structure evaluation reports, reducing variability in reporting formats. While ownership and adoption remain asymmetric, those evaluation managers and country offices that systematically apply these tools have demonstrated improvements in the coherence and completeness of the evaluation process and related outputs.
- The involvement of RPMEOs in reviewing ToRs and key deliverables has strengthened quality control where consistently applied, despite the fact that their responsibilities

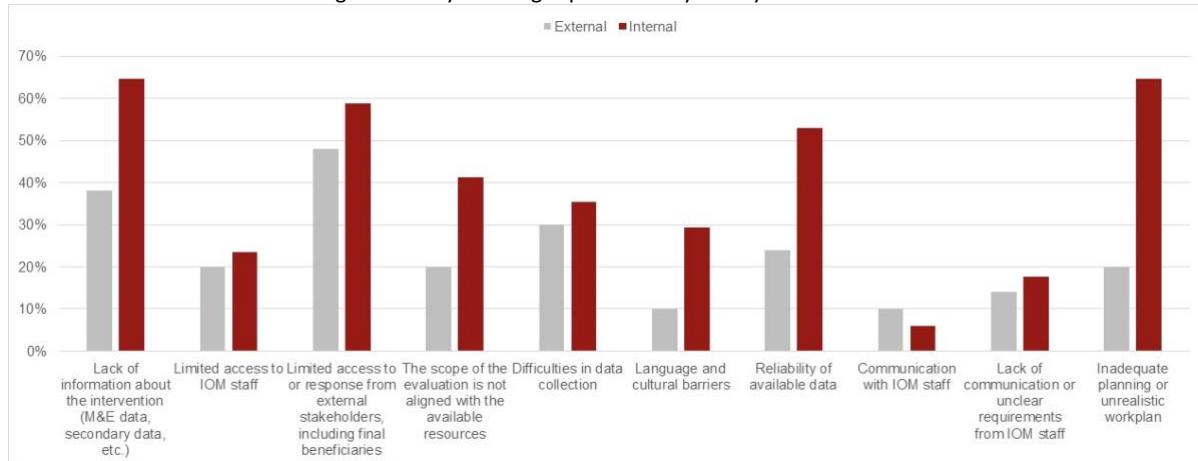
have increased since 2023. Additionally, structured technical support for evaluation managers, especially in regions where internal review mechanisms are more developed, has improved the methodological standards of some evaluations and the quality and dissemination of their reports. Some regional offices and RPMEOs have developed tools, strategies and procedures that could be shared with – and would benefit – other peers.

- Hands-on technical guidance and ad hoc capacity-building efforts from the regional offices, especially those providing step-by-step methodological support, have contributed to better-structured evaluations. Internal evaluators and managers who receive sustained technical guidance throughout the evaluation cycle tend to produce reports with greater analytical depth and stronger alignment with institutional standards. However, these initiatives remain sporadic and unevenly distributed across regions.
- The introduction of standardized indicators, the Strategic Results Framework and the promotion of improved monitoring mechanisms (for example, concerning the quality of the results matrices and the related measurement tools) have strengthened the link between evaluations and institutional priorities. In some cases, this alignment has resulted in more structured assessments of programme outcomes, improving the potential for evaluations to inform future programming and policy decisions. Considering that the changes in the design of results matrices, formulation of indicators and reporting have been introduced recently, the positive results will likely be more visible in the next evaluation cycles.

However, while quality assurance tools, standardized templates and specialized training materials have helped improve consistency in evaluation processes, areas still need improvement. Technical, methodological and operational limitations (such as the poor adaptation of some evaluations to practical needs on the ground, the limited quality of some results matrices and monitoring systems or the absence of decentralized and clear quality control procedures) challenge the progress achieved in recent years.

The following figure compares the challenges faced by external and internal evaluators. It highlights the practical difficulties unique to each and the common structural and technical constraints that hinder the evaluation process.

Figure 24. Key challenges perceived by surveyed evaluators



Lack of information about the intervention posed a challenge for a dramatically greater proportion of internal evaluators (65%) than for external evaluators (38%). Inadequate planning or unrealistic workplans presented an equally significant concern for internal evaluators (65%), while only 20% of external evaluators perceived this to be a challenge.

Limited access to external stakeholders and final beneficiaries affected a majority of internal evaluators (59%) and nearly half of external evaluators (48%). The reliability of available data proved problematic for 53% of internal evaluators compared to just 24% of external evaluators.

Resource-scope misalignment created difficulties for 65% of internal evaluators but affected only 20% of external evaluators. Data collection challenges were reported by 35% of internal evaluators and 30% of external evaluators. Language and cultural barriers were mainly a challenge for 29% of internal evaluators identifying these as a challenge.

Communication with IOM staff represented the only challenge more commonly reported by external evaluators (10%) than internal evaluators (6%). Overall, internal evaluators consistently reported higher rates of challenge across nearly all evaluation areas compared to external evaluators.

A significant number of the evaluation managers interviewed also pointed out two additional structural challenges that are worth mentioning:

- High staff turnover and technical expertise gaps significantly impact the quality of evaluations. The lack of sustained capacity-building programmes and reliance on ad hoc training have led to evaluation quality depending more on individual expertise than institutionalized knowledge. In some cases, evaluations were conducted by individuals lacking specialized methodological training, which resulted in inconsistencies in approach and analytical weaknesses. Despite progress in training over time, its intensity has declined significantly since 2021. Coupled with ongoing staff turnover, this has created structural challenges for IOM in retaining knowledge and maintaining best practices.
- Structural weaknesses in evaluation planning have led to rushed processes and limited analytical depth. Short evaluation time frames and delays in commissioning evaluations

constrain the ability to conduct thorough data collection and analysis. This often results in last-minute adjustments, weak methodological choices and reports prioritizing compliance over substantive learning. Though organizational culture has improved regarding the importance of evaluations, a widespread perception remains that they are a contractual requirement with no direct or clear added value in learning, management, or specialization.

The main findings regarding evaluation design and implementation processes are presented below. The interviews with IOM staff and evaluators, as well as the open questions included in the surveys, provided accurate information on each stage, identifying the main strengths and weaknesses. The following two sections summarize the most relevant ideas provided by key informants.

#### 4.1 Evaluation design

Institutional policies, donor expectations and operational considerations guide the design of evaluations within IOM.<sup>38</sup> However, evaluations and their design remain highly decentralized, leading to some inconsistencies in implementation. Although evaluations are widely acknowledged as essential for accountability, learning and strategic decision-making, their planning often lacks a systematic and enforceable framework. Interviewees highlighted that evaluations are frequently conducted as routine exercises rather than being prioritized where they can generate the most value. Without a clear and precise enforcement mechanism, the decision on which projects to evaluate and what specifically to evaluate is often left to PMs, whose priorities vary across regions, types of interventions and thematic areas.

The definition, characteristics and timing of evaluations are influenced by institutional policies and donor requirements; however, their design and methodological development often lack strategic coherence. Evaluations are essential for large-scale, multi-year, or high-impact projects, particularly those testing innovative approaches or representing pilot initiatives. Within IOM, evaluations are most often conducted for large-scale projects lasting one year or more. In contrast, short-term interventions and those with limited budgets are seldom evaluated unless they contribute to broader institutional learning objectives. Although training initiatives since 2018 have encouraged technical teams to incorporate evaluations into proposals, this practice has not been systematically adopted. The revised version of the yet-to-be-approved evaluation policy (2025) aims to establish a more specific budget threshold for mandatory evaluations, but so far compliance remains inconsistent without an enforcement mechanism.

Key informants indicated that evaluation timing is another critical concern, with many evaluations being initiated too late in the project cycle, often as a last-minute response to donor reporting requirements rather than as a strategic and well-planned learning exercise. In some cases, evaluations begin less than two or three months before project closure, limiting the time available for comprehensive data collection, stakeholder engagement and rigorous analysis. This reactive approach, driven more by administrative deadlines than by internal

---

<sup>38</sup> Guidance on evaluations can be found in all of the following IOM resources: IOM, *IOM Project Handbook* (Geneva, IOM, 2017); IOM, *IOM Monitoring and Evaluation Guidelines* (Geneva, IOM, 2020); IOM Central Evaluation Unit, “*Guidance for evaluation managers*” (November 2021); IOM Central Evaluation Unit, “*Guidance on quality management of IOM evaluations*” (September 2022); IOM Central Evaluation Unit, “*Guidance on the use of evaluations and follow-up of evaluation recommendations*” (September 2022); IOM Development Fund, “*Evaluation guidelines for IOM Development Fund projects*” (15 January 2025).

planning frameworks, undermines the capacity of evaluations to inform real-time decision-making and adaptive project management.

Financial considerations, project complexity and the need for perceived objectivity have influenced the choice between internal and external evaluations. Internal evaluations have traditionally been cost effective, particularly for smaller projects, utilizing IOM staff expertise and institutional knowledge. However, their credibility is often questioned, raising concerns about potential bias. Additionally, internal evaluations are constrained by competing responsibilities, as evaluators must balance their evaluation duties with other operational tasks. A structured internal evaluation model that involved trained IOM staff from outside the country office was initially considered a promising alternative. However, it has proven difficult to sustain due to logistical constraints and a lack of managerial support, as concluded by the UNEG/OECD peer review.<sup>39</sup> Conversely, external evaluations are preferred for their independence and methodological rigour, especially for large-scale or complex projects. However, interviewees noted that the quality of external evaluations can vary significantly, often depending more on the evaluator's expertise than on their external status. The criteria for selecting external evaluators are sometimes inadequate, focusing primarily on the requirement for externality and independence rather than ensuring methodological and thematic competence. The shift away from internal evaluations results in situations where projects struggle to identify viable evaluation options, particularly where donor funding for evaluations is limited.

The elaboration of ToRs is a crucial phase in evaluation design, but the process is decentralized and often lacks methodological rigour. PMs are responsible for drafting ToRs, but their focus is frequently limited to identifying key stakeholders rather than critically defining evaluation criteria and questions. Many ToRs include all OECD-DAC criteria without assessing their specific relevance to the project, leading to broad and unfocused evaluation scopes. While RPMEOs and Headquarters can provide guidance and quality control, their involvement is inconsistent and depends largely on whether PMs actively seek their support. Compliance remains voluntary despite a 2022 guidance document recommending a systematic review of the ToR by RPMEOs. As noted in interviews, many PMs do not proactively submit ToRs for review. This lack of structured oversight results in significant variations in evaluation clarity, methodological soundness and strategic focus. Efforts such as webinars and microlearning sessions were introduced to improve ToR quality, but their integration into evaluation planning is inconsistent, and their implementation remains discontinuous. This further contributes to disparities in evaluation design.

The quality assurance tools for ToRs, inception reports and final reports have been introduced to enhance evaluation standards. However, their application is still inconsistent across IOM country offices and technical teams. Based on the original meta-evaluation scoring tool developed for the 2017–2019 assessment, these tools align with international standards, such as those from the OECD, ALNAP and UNEG, providing a structured approach to maintaining evaluation quality. However, many evaluation managers and programme staff struggle to integrate them effectively into their workflows. Some of the focal points within the regional offices considered that a key challenge is the lack of awareness and accessibility, with many country offices either unaware of these tools or not perceiving them as mandatory. While the

---

<sup>39</sup> UNEG, “[UNEG/OECD peer review of the International Organization for Migration \(IOM\) evaluation function](#)” (May 2021).

introduction of quality assurance tools has improved adherence to standardized templates, their impact is limited by the absence of structured training and follow-up mechanisms. The tools are considered lengthy and highly technical, making them difficult for staff with limited evaluation expertise to navigate. As a result, some evaluators apply them selectively, while others bypass them altogether. Furthermore, the discontinuous nature of the training programmes has not allowed for the consolidation of knowledge in all country offices, which are often affected by constant staff turnover. Interviewees suggested simplifying materials, introducing visual summaries, and providing more hands-on coaching and mentoring rather than relying solely on written guidance. Additionally, regional offices have tried improving accessibility by creating SharePoint folders with key resources. Still, these initiatives remain informal and depend on individual rather than institutionalized processes. Without clearer enforcement measures, such as integrating quality assurance checks into mandatory evaluation workflows or requiring quality assurance tool application, their use will probably remain inconsistent.

In summary, while policies and tools exist to support evaluation planning, their application remains highly variable, and evaluation design still reflects some structural gaps that influence their overall effectiveness. Strengthening institutional oversight, formalizing the review process for ToRs, ensuring the systematic integration of evaluations into project planning and making quality assurance tools more accessible and enforceable are critical steps towards improved evaluation design.

### 1.2 Evaluation implementation

The implementation of evaluations within IOM is influenced by a combination of institutional, operational and procedural factors that affect consistency and quality. Implementation often relies on PMs' initiative, technical support availability, and the degree of engagement from RPMEOs. In some instances, evaluations are well structured when they benefit from proactive planning and strong quality assurance; however, in many cases, irregular supervision, inconsistent methodological oversight and limited institutional coordination impede the process. The potential support role of RPMEOs during the evaluation process is not fully utilized, while structural challenges – such as resource limitations, compressed timelines and the lack of formal mechanisms for revision and feedback – also complicate the process. Achieving a more harmonized and robust implementation requires clearer institutional mechanisms that ensure evaluations are systematically carried out, methodologically sound and strategically useful.

A key consideration is that consultation and supervision throughout the evaluation process vary significantly among country offices, affecting quality and methodological alignment. While institutional guidance expects programme managers and RPMEOs to provide technical input, their engagement level is inconsistent and largely depends on individual initiative. In some cases, PMs actively involve regional offices in refining methodology and sources of data, while in others, evaluations proceed with minimal or no external oversight, resulting in gaps in quality control. IOM staff play an active role in logistical aspects – such as identifying informants, facilitating stakeholder engagement and providing access to documentation – but their involvement in defining evaluation criteria or ensuring methodological soundness is often limited. This is particularly evident in country offices without dedicated M&E staff, where programme managers take on evaluation responsibilities despite lacking technical expertise.

Without a formal requirement for regional oversight at key stages, the involvement of RPMEOs remains largely reactive rather than an integral part of the process.

Likewise, supervision across different phases of an evaluation remains highly uneven due to the absence of structured quality control mechanisms. While experienced evaluation managers ensure rigorous oversight by monitoring deliverables and providing iterative feedback, others take a more hands-off approach, finalizing evaluations without thorough internal review. The selection of key informants and reference documents follows a similarly unstructured pattern. Informant lists are often compiled based on convenience rather than systematic sampling, increasing the risk of selection bias and limiting the diversity of perspectives captured. The document selection is also of varied quality: while some evaluations benefit from a broad range of sources, others rely predominantly on internal reports, overlooking external assessments that could provide a more objective view. Without clear institutional guidelines, the credibility and comprehensiveness of evaluation findings are affected.

When challenges arise during evaluations, responses are often swift and effective. However, they are typically ad hoc and personalized, lacking guidance from established institutional procedures or a set code of good practices. Some challenges identified during the interviews include tight deadlines, insufficient budgets, limited stakeholder access and weak methodological execution. However, how these issues are addressed primarily depends on the experience and initiative of the evaluation manager. More experienced staff anticipate challenges and proactively engage regional offices for technical support, while those with less expertise may struggle to identify and resolve methodological weaknesses. Additionally, the absence of a standardized mechanism for escalating difficulties or ensuring that corrective actions are documented makes it challenging to track recurring issues and institutionalize lessons learned. The decentralized nature of the IOM evaluation system further complicates problem solving, as there is no formal requirement for country offices to seek support or report significant implementation challenges. Consequently, evaluations frequently attempt to integrate some of these conditions independently without introducing major changes, leaving limited opportunities for mid-course adjustments or methodological refinements.

Another relevant point is that the connection between monitoring and evaluation has improved in this most recent period (2020–2024) but remains limited. This limitation hinders the effective use of monitoring data to inform evaluations. While structured monitoring frameworks (for instance, more robust results matrices, standardized indicators, PRIMA reports) have enhanced documentation and analysis, there is little integration between ongoing project monitoring and evaluation design. Many evaluations still suffer from methodological gaps and lack structured mechanisms to ensure that monitoring data informs evaluation priorities, methodologies, or findings. Additionally, country offices do not have a clear idea or approach regarding whether and how evaluations should assess progress against the Strategic Results Framework indicators,<sup>40</sup> which reduces their potential to generate insights and impactful reports.

---

<sup>40</sup> The Strategic Results Framework indicators are standardized metrics used by IOM staff during the design of results matrices to systematically monitor, measure and report on progress towards achieving institutional goals, allowing a better alignment between project outcomes and programming documents.

Despite these challenges, several evaluation practices have proven effective in enhancing planning, ensuring quality assurance and boosting the overall credibility of findings. PRIMA and other customized tools have increased transparency regarding evaluation timelines and compliance. Additionally, some regional offices have implemented quarterly trackers and direct follow-ups to encourage better preparation. According to interviewees, evaluations tend to be more structured and methodologically sound when regional offices are involved early in the evaluation process, particularly in drafting ToRs and conducting quality checks. Strategic midterm evaluations have also shown value by facilitating real-time learning and adjustments, rather than limiting evaluations to retrospective assessments. Furthermore, the training materials and guidelines available to evaluation managers have strengthened the evaluation culture within IOM, raising awareness of the importance of these analyses and reports for prospective planning, strategic communication and the accumulation of know-how. The use of standardized templates and quality assurance tools has bolstered methodological rigour, reducing evaluation inconsistencies.

## Findings on evaluation capacities and skills

Since 2020, IOM has made significant strides in enhancing the quality of its evaluations. Recognizing the essential role evaluations play in strengthening accountability, learning and evidence-based decision-making, the Organization has concentrated on developing and updating evaluation guidelines while building staff capacity. The following section highlights the key findings in this area, based on key informant interviews and online survey results.

### Key findings

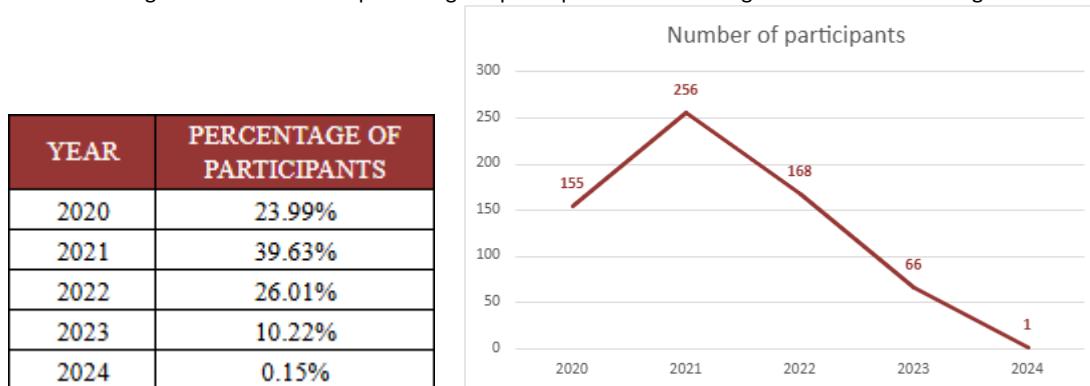
19. The internal evaluation training in monitoring and evaluation, which peaked significantly in 2021 with 256 participants, has drastically declined since 2022, reaching only 1 participant in 2024. This sharp reduction may be attributed to factors such as staff turnover, strategic shifts within the Organization, or decreased demand for training. Additionally, budget constraints, shifting institutional priorities and a possible saturation effect – where a significant portion of relevant staff had already been trained – may have further contributed to this decline.
20. The most consulted evaluation-related documents are the *IOM Project Handbook* and the *IOM Monitoring and Evaluation Guidelines*. However, the inconsistent use of other key documents, such as the “Guidance for addressing gender in evaluations”, might contribute to the lack of inclusion of cross-cutting issues.
21. Evaluation managers and internal evaluators perceive themselves as skilled, motivated and committed to managing and conducting evaluations. However, evaluation managers rated themselves less highly, with regard to all three characteristics, than internal evaluators rated themselves. While evaluation managers reported perception less supported than internal evaluators, they expressed a higher level of engagement in influencing others in evaluation-related matters.

#### 5.1. Internal evaluator training

The IOM M&E strategy of 2018–2020 laid the foundation for the Organization’s training on monitoring and evaluation. This strategy led to the launch of an e-learning course that trained over 194 staff members by 2020 and the establishment of an internal evaluator programme. The IOM *OIG Monitoring and Evaluation Strategy 2021–2023* continued these efforts, aiming to strengthen the evaluation function across the Organization. This strategy sought to create a strong culture of evaluation, enhance evaluation quality, ensure stable financial support and further develop staff skills by integrating M&E competencies into various roles within the Organization. The CED “*Evaluation strategy 2024–2028*” builds upon these efforts, aiming to continue strengthening the evaluation capacity and culture. This is expected to be achieved through the update of the I-learn and E-campus training courses, providing the tools and systems to support effective training, and offering support for training whenever necessary.

From 2020 to 2024, 646 individuals received training (see Figure 25). The highest number of participants was in 2021, with 256 people. This surge was likely due to a strong focus on capacity-building linked to the OIG M&E strategy 2021–2023, which prioritized enhancing M&E knowledge and skills. However, participation dropped significantly in 2022 to 168, a 66 per cent decrease from 2021. The downward trend continued in 2023, with only 66 participants, and declined again in 2024, with just 1 participant.

Figure 25. Number and percentage of participants in monitoring and evaluation training



Although details are lacking, the decline in M&E training participation after 2021 may have resulted from staff turnover, strategic shifts with budget constraints and saturation. With many trained in 2020–2021, fewer new staff required training, reducing demand. Some key informants see the end of these trainings as a significant loss.

In response to the survey question about participation in evaluation training, 15 out of 36 evaluation managers and 11 out of 17 internal evaluators reported receiving training from IOM.<sup>41</sup> This means 65 per cent of evaluators and 42 per cent of evaluation managers benefited from training. However, gaps remain, highlighting the need to evaluate IOM training accessibility, reach and effectiveness to support both groups better.

## 5.2. Use of key evaluation documents

A key strategy in the Organization’s endeavour to improve the quality of its evaluations has been developing and revising evaluation guidelines, handbooks and templates. Some

<sup>41</sup> The training programme with the highest participation (4 attendants) among evaluation managers was the “Project development, project management, and the M&E e-learning course”. The training programme with the highest participation among evaluators (5 attendants) was the “M&E e-learning course, internal evaluator training”.

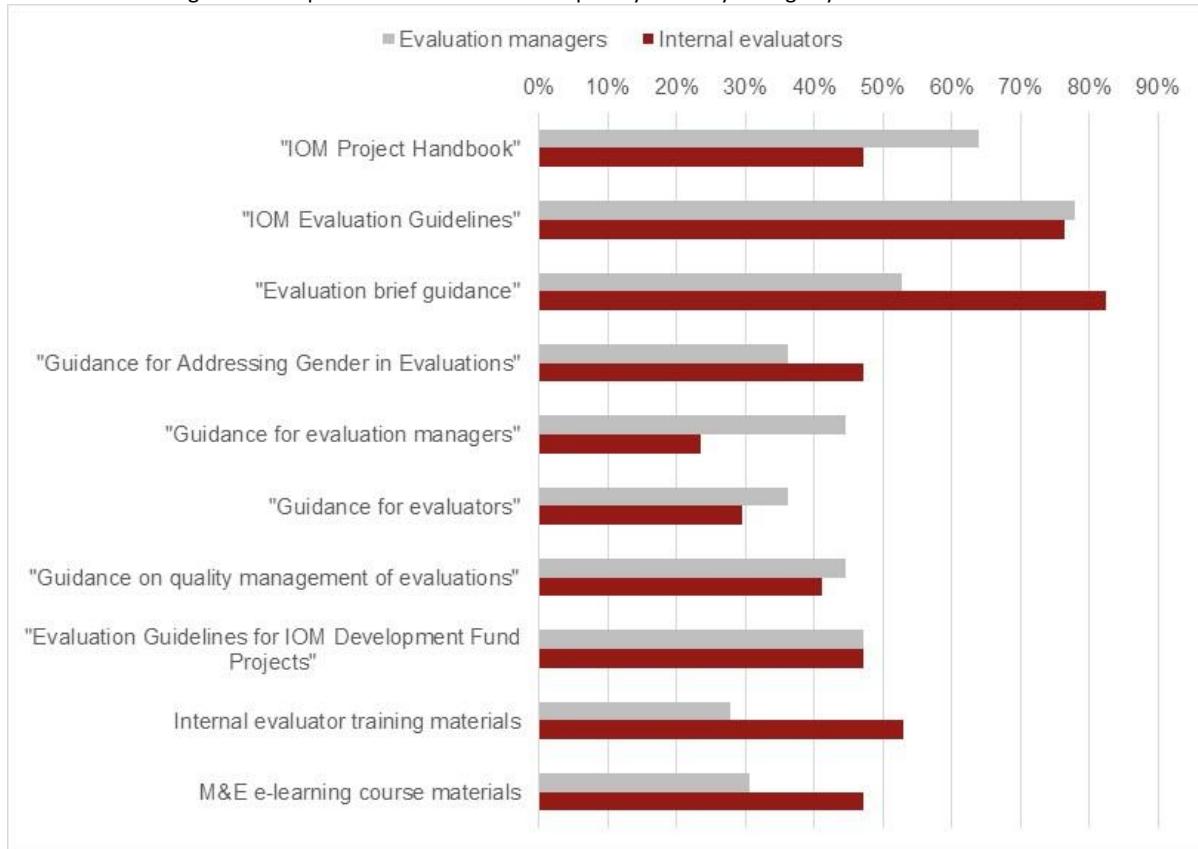
publications include the *IOM Monitoring and Evaluation Guidelines* (published 2021, [digital version](#) available since 2023), “Guidance for evaluation managers” (2021), “Guidance on quality management of IOM evaluations” (2022), “Guidance on the use of evaluations and follow-up of evaluation recommendations” (2022) and the “Evaluation guidelines for IOM Development Fund projects” (2025).

The *IOM Monitoring and Evaluation Guidelines* (2021) and the *IOM Project Handbook* (2017) are the resources most commonly consulted by evaluation managers. However, the IOM “Guidance for addressing gender in evaluations” (2018) is less frequently referenced, aside from training-related materials. This gap could potentially impact gender mainstreaming in evaluations, as highlighted by the quality assessment of cross-cutting issues. It was observed that most evaluation methodologies exclude techniques for capturing information related to cross-cutting questions. Consequently, many reports only present isolated findings on gender, and conclusions often lack a gender perspective.

During the interviews with M&E officers, it was highlighted that the primary issue is not the absence of quality assurance tools, but their lack of depth and practical guidance. Despite their quality, evaluation materials are often difficult to apply due to their dense, text-heavy format, which makes them less accessible to staff with limited evaluation expertise or time constraints. The main barriers identified included a lack of time to study lengthy documents, difficulty understanding technical content, competing priorities within project management roles and language barriers for non-English-speaking staff. Existing checklists provide a basic framework but fail to offer concrete guidance on selecting relevant evaluation criteria, defining appropriate methodologies, or ensuring coherence between findings and recommendations. Many programme managers and M&E officers lack formal evaluation training, which makes applying quality assurance tools effectively challenging. Some regional offices provide ad hoc guidance to address this gap, but it is not standardized across the Organization. Furthermore, there is no enforced institutional mechanism to ensure the consistent application of quality assurance tools, leading to evaluations often proceeding without proper regional review and lacking a systematic quality control process before finalizing reports.

For internal evaluators, the *IOM Monitoring and Evaluation Guidelines* and the “[Evaluation brief guidance](#)” (2018) are the most frequently used resources, while the “Guidance for evaluators” (2021) is consulted less often. Apart from these, the rest of the guides are consulted by at least 40 per cent of respondents, indicating widespread use among internal evaluators. Figure 26 illustrates the identified patterns.

Figure 26. Proportion of stakeholders frequently or always using key evaluation documents



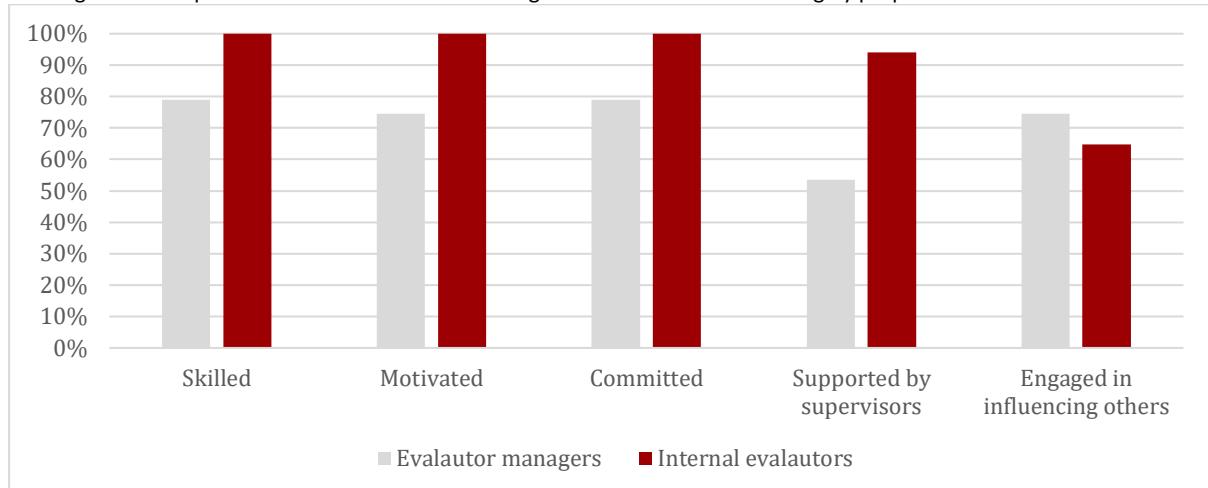
### 5.3. Self-perceived preparedness of internal evaluators

This subsection analyses the self-perceptions of evaluation managers and internal evaluators regarding their readiness to manage and conduct evaluations. The information is based on survey results from a specific question in which respondents rated their level of preparedness across five categories: skills, motivation, commitment, support provided by their unit or supervisor and engagement in influencing other colleagues to achieve evaluative excellence.

As shown in figure 27, both evaluation managers and internal evaluators perceive themselves as skilled, motivated and committed to managing and conducting evaluations. However, evaluation managers rated themselves lower than did internal evaluators across all three categories. This discrepancy could be linked to their lower participation in evaluation-related training, which may impact their confidence in their evaluation skills.

Regarding the support received from supervisors, evaluation managers reported perception less well supported than did internal evaluators. However, they expressed a higher level of engagement in influencing others in evaluation-related matters. This suggests that despite perception less supported, evaluation managers actively guide internal evaluators and promote the use of evaluation findings within the Organization. Their engagement in influencing others could reflect their commitment to fostering an evaluation culture and ensuring that evaluation processes contribute to decision-making and organizational learning.

Figure 27. Proportion of stakeholders considering themselves somewhat or highly prepared to conduct evaluations



## Conclusions

### Evaluation reports

#### Conclusion 1

The trajectory of the quality of IOM evaluations is marked by a peak in 2020, followed by a substantial decline to its lowest point in 2022 and a subsequent moderate recovery in 2023.<sup>42</sup> This decline aligns with the impact of COVID-19, which disrupted project objectives, limited access to key informants and hindered evaluation coordination. Assessed reports indicate extended deadlines and modified methodologies, often without adequate alternatives. The pandemic likely delayed the benefits of using the guidelines launched in 2022 as methodological challenges limited their implementation. While these guidelines set a framework for improvement, their impact on report quality remains unclear due to the limited use. With signs of recovery post-2022, sustaining and enhancing evaluation standards will require addressing methodological gaps, strengthening adaptive strategies and ensuring consistent quality assurance.

Conclusion 1 is linked to findings 1, 2, 8 and 9.

#### Conclusion 2

While evaluations generally “approach requirements” and certain components exhibit high quality, there is a critical need to address the gaps identified, particularly in integrating cross-cutting perspectives, refining methodologies and knowledge management, especially lessons learned and good practices. Although they meet the minimum quality standards, it is necessary to continue paying attention to the quality of the project background, recommendations, and executive summary. Enhancing these areas will improve the quality of evaluation reports and bolster their utility and impact in informing decision-making and fostering continuous improvement.

Conclusion 2 is linked to findings 1,3, 4, 5 and 6.

<sup>42</sup> The preliminary score for 2024 remains uncertain due to the limited number of reports available for assessment; thus, 2024 should be included in the next meta-evaluation.

## **Evaluation design**

### **Conclusion 3**

Regional offices have played a crucial role in enhancing evaluation quality by providing practical technical guidance and on-the-spot capacity-building initiatives, particularly through comprehensive methodological support. The early involvement of regional offices in drafting ToRs and strategic midterm evaluations have facilitated real-time learning and adjustments. Furthermore, evaluation quality is highest where regional offices and RPMEOs are actively involved in the review of ToRs and key deliverables, suggesting that institutional support is a key determinant of methodological rigour. Evaluation reports often demonstrate deeper analysis and stronger adherence to institutional standards when internal evaluators and managers are consistently guided through the evaluation process.

Conclusion 3 is linked to findings 12 and 16.

### **Conclusion 4**

The inadequate time allocation for the evaluation process has led to tight timelines and superficial analysis, limiting the depth and rigour of assessments. Delays in commissioning evaluations and short time frames have hindered the data collection and analysis processes, often forcing last-minute adjustments and methodological compromises. As a result, many reports prioritize compliance to donor requests over meaningful learning (for example, evaluation questions are not adapted to institutional or sectoral needs, evaluations are launched shortly before the end of implementation and some PMs conceive of them as pure accountability exercises). Although the Organization's evaluation culture has improved since 2020, evaluations are still largely seen as a contractual obligation rather than an integral tool for learning and decision-making.

Conclusion 4 is linked to finding 12 and 15.

### **Conclusion 5**

The proportion of the total budget dedicated to evaluation significantly influences evaluation quality. This underscores the need to dedicate an appropriate percentage of a project's budget to evaluation, rather than focusing on absolute amounts. Despite the structural adjustments following reductions in donor funding, it is crucial to ensure in the initial formulation that a dedicated percentage of the project budget is set aside for evaluation. This approach ensures enough resources are available for thorough evaluations, leading to valuable insights and actionable recommendations that foster organizational improvement.

Conclusion 5 is linked to findings 7 and 12.

## **Evaluation implementation**

### **Conclusion 6**

Despite the challenges in ensuring methodological rigour and consistent quality assurance, several effective practices have enhanced evaluation processes within IOM. Tools like PRIMA and proactive measures by regional offices, such as quarterly trackers and direct follow-ups, have increased transparency and improved preparation, resulting in more structured evaluations. Comprehensive training materials have strengthened the evaluation capacities

and culture of the Organization. The use of standardized templates and quality assurance tools has reduced inconsistencies, thus increasing the credibility of findings. Some regional offices and RPMEOs have also developed effective tools, strategies and procedures that could be shared to benefit broader institutional learning. To maximize impact, there is a need for clearer institutional mechanisms that empower M&E officers, ensure systematic evaluations and align methodologically sound practices with strategic goals. By building on these effective practices and addressing existing gaps, IOM can further elevate the quality and impact of its evaluations to support strategic decision-making and organizational advancement.

Conclusion 6 is linked to findings 17 and 18.

### **Conclusion 7**

The introduction of standardized indicators (the Strategic Results Framework) and the push for better monitoring mechanisms have helped to reinforce the connection between evaluations and institutional priorities. This stronger alignment has occasionally led to more organized assessments of programme outcomes, thereby boosting the evaluations' ability to guide future programming and policy decisions. Yet, given that adjustments to results matrices, indicator formulation and reporting are relatively new, the complete effects of these changes are anticipated to be clearer in upcoming meta-evaluation cycles.

Conclusion 7 is linked to findings 15 and 16.

### **Conclusion 8**

Decentralization without standardized and clear procedures for design, approval and quality control has led to inconsistent evaluation quality, often dependent on the experience and engagement of individual PMs. The absence of systematic oversight, reliance on voluntary compliance and unclear roles and responsibilities further weaken quality assurance. RPMEOs are underutilized in evaluation processes and disproportionately tasked with strategic planning and monitoring; the challenges they face are compounded by structural limitations, scarce resources and tight timelines. Institutional mechanisms for empowering M&E officers remain vague, hindering consistent evaluation practices and alignment with IOM strategic goals. Strengthening these mechanisms would significantly improve evaluation quality and utility while fostering a culture of learning, adaptation and continuous improvement across the Organization.

Conclusion 8 is linked to findings 13 and 17.

### **Conclusion 9**

Inconsistent promotion of quality assurance mechanisms and a lack of methodological rigour undermine the credibility and depth of evaluations. Weak sampling strategies, insufficient triangulation and inadequate traceability and transparency in data collection and analysis have led to often superficial findings and have failed to provide robust evidence for conclusions and recommendations. The disconnect between findings and recommendations reduces evaluations' strategic value. Many reports fail to establish a clear and logical link between analysis and proposed actions, leading to recommendations that are too broad and not evidence based. The absence of structured stakeholder engagement and participatory discussion in the recommendation process further weakens their relevance and applicability.

Conclusion 9 is linked to findings 14 and 16.

### **Terms of reference**

#### **Conclusion 10**

Well-crafted ToRs are positively correlated with more robust evaluation reports. Although the “Guidance on quality management of IOM evaluations” (2022) lays a foundation for improving the quality of ToRs, many ToRs continue to exhibit inconsistencies and gaps in clarity, focus and methodological rigour. The blanket inclusion of all OECD-DAC criteria and the lack of contextual adaptation leads to broad scopes and diluted analytical depth. Structural factors such as the expertise of the evaluation team and managerial support also influence outcomes. Regional offices, often reliant on the initiative of specific PMs, coupled with high staff turnover and limited technical capacity in some country offices, further undermine consistency. These factors contribute to varying levels of clarity, focus and methodological rigour within ToRs, ultimately weakening the overall quality of evaluations. Addressing these structural gaps is essential to sustaining and enhancing the quality of ToR and, consequently, the overall effectiveness of evaluations.

Conclusion 10 is linked to findings 10 and 13.

### **Quality assurance tools**

#### **Conclusion 11**

The assessment of the usefulness of IOM methodological guidelines and quality assurance tools reveals a varied and complex reality across different stakeholders. This suggests that their effectiveness and application differ based on context, user familiarity and specific evaluation needs, rather than presenting a uniform or consistently applied approach. These guidelines and tools provide a robust framework that evaluation managers appreciate for its structured approach. However, their complexity, length and inconsistent adoption pose significant challenges in practical application, undermining their effectiveness. Internal evaluators often view these guidelines as overly rigid, which can lead to their mechanical application rather than thoughtful adaptation to specific evaluation contexts. Some project and evaluation managers struggle to use them due to their complexity, length and technical nature, while others are unaware of their existence or do not perceive them as mandatory. The discontinuity of training, frequent staff turnover and lack of structured follow-up mechanisms hinder their consistent and effective use. This divergence in stakeholder perspectives underscores the pressing need for a more balanced approach that can accommodate both methodological rigour and the flexibility necessary to address the diverse requirements of evaluations.

Conclusion 11 is linked to findings 14 and 15.

#### **Conclusion 12**

The inconsistent uptake of quality assurance tools signals that voluntary compliance is insufficient. Embedding quality assurance checks into mandatory approval workflow promotes uniformity, reliability and methodological rigour. Requiring the use of quality assurance tools as a condition for evaluation approval helps guarantee that all evaluations meet established quality benchmarks before moving forward. This proactive approach not only enhances the

quality and credibility of evaluation outputs but also fosters a culture of accountability and continuous improvement. In the absence of such structured enforcement, quality assurance tool usage is likely to remain irregular, increasing the risk of inconsistency, inefficiency and reduced organizational effectiveness.

Conclusion 12 is linked to finding 17.

### **Capacities and skills**

#### **Conclusion 13**

Persistent gaps in technical capacity – exacerbated by high staff turnover and sporadic training – highlight a systemic challenge in institutionalizing evaluation skills and in applying best practices. Building a resilient evaluation function will require long-term investment in structured learning pathways, rather than ad hoc efforts.

Conclusion 13 is linked to finding 17, 19, 20 and 21.

### **Good practices**

Some evaluation processes and strategies within the Organization have demonstrated effective approaches that enhance methodological rigour and contribute to the standardization of evaluation outputs. The following good practices have been identified, showcasing significant potential for scalability and methodological development.

The active involvement of RPMEOs has played a crucial role in enhancing evaluation. In offices where RPMEOs have been actively involved, evaluations have shown better methodological rigour and coherence. Their oversight, particularly in reviewing ToRs, methodologies and final reports, has led to stronger evaluation reports with clearer recommendations. However, the effectiveness of this role varies widely across regions, as some country offices do not systematically involve RPMEOs in the evaluation process.

Developing internal evaluation databases and knowledge management platforms has improved access to previous evaluations, offering specific references for formulating new ToRs and methodological approaches. These platforms enable IOM staff to leverage past findings to inform new projects, allowing technical teams to integrate previous evaluation insights into current programming. However, despite their existence, usability and accessibility remain challenging, as not all staff are aware of these resources or how to utilize them effectively.

Adopting standardized evaluation formats, particularly through structured templates and checklists as used in the IOM Development Fund evaluations, has enhanced consistency in reporting. Most evaluations now adhere to predefined formats, streamlining the documentation process and ensuring the inclusion of key components such as methodology, findings and recommendations. While this standardization does not inherently guarantee content quality, it has significantly improved the coherence and comparability of evaluation reports across different offices.

## Lessons learned

Despite the notable advancements outlined in the previous sections, several structural challenges continue to hinder the effectiveness of evaluations. The following points highlight systemic issues that impact evaluations' quality, depth and strategic use, emphasizing key areas that require further attention.

The insufficient integration between monitoring and evaluation functions has created challenges in leveraging robust data and evidence to inform evaluations, thereby limiting the depth and reliability of their findings. Many assessments rely on weak or incomplete monitoring data, often due to poorly designed results matrices and limited indicators that fail to capture key outcomes. In multiple cases, project monitoring indicators have been overly broad, lacking specificity and clear measurement criteria, making it difficult for evaluations to produce conclusive insights. Although monitoring is intended to function as an ongoing learning tool that informs evaluations, in practice, the two functions still operate separately in many offices.

Limited technical expertise among some evaluation managers and programme staff responsible for overseeing evaluations has negatively impacted the methodological rigour of many reports. In several instances, evaluations have been conducted by individuals without formal training in evaluation methodologies, leading to inconsistencies in data collection, weak sampling strategies and a lack of analytical depth in findings. This issue seems to be particularly pronounced in country offices without dedicated M&E staff, where programme managers with limited evaluation experience have had to lead the process.

The inconsistency in institutional and donor requirements for evaluations have also impacted their prioritization within the Organization. Unlike financial audits, which are mandatory for most projects, evaluations are not systematically required by donors. As a result, some projects proceed without any evaluation component, while others include evaluations as a formality rather than a tool for learning and accountability. The lack of external pressure to conduct evaluations and an unclear internal policy have contributed to an organizational culture where evaluations are often perceived as optional activities rather than an essential part of the project management cycle.

## Recommendations

### Recommendations to enhance evaluation report quality

#### Recommendation 1

Keep promoting the use of the evaluation report templates and quality assurance tools, and ensure that it is mandatory to have all sections accordingly developed. Evaluation managers and RPMEOs should:

- Consider and promote the inclusion of the following sections during the evaluation drafting phase:
  - Executive summary. Ensure that conclusions, lessons learned, and good practices are included when relevant, to make this section comprehensive and useful.

- Project background. Include details on the project budget. Expand on identifying and characterizing stakeholders, with a special focus on rights holders.
- Evaluation background. Clearly define the evaluation users and how the results of the evaluation can be used in decision-making.
- Criteria and evaluation questions. Promote the development of high-quality evaluation matrices to guide the process effectively. Ensure that all cross-cutting themes are consistently integrated throughout all criteria.
- Evaluation methodology. Clearly distinguish methodological approaches from quantitative and qualitative methods, as the methodological approach serves as the guiding framework for the evaluation, shaping its overall design. Provide greater detail on data gathering and data assessment techniques, ensuring effective data triangulation, enhanced sampling processes, and stronger stakeholder participation.
- Findings. Deepen the analysis by integrating data gathered using multiple methods, ensuring a comprehensive perspective. Address all cross-cutting themes consistently and emphasize unexpected findings, regardless of their nature.
- Conclusions. A clear distinction should be maintained between findings and conclusions, with value judgement regarding the project performance reserved for the conclusions section.
- Cross-cutting themes should be fully integrated into the conclusions, as they should be throughout all sections of the report.
- Recommendations. Enhance the practical applicability of recommendations by clearly specifying the resources needed for implementation and establishing their priority level.
- Lessons learned and good practices. The inclusion of these sections should be explicitly defined in the ToRs, reinforcing their role as essential components of evaluation reports. Both sections require improved description, contextualization and practical applicability to ensure effective implementation.
- Ensure adherence to IOM and UNEG ethical principles and standards, explicitly stating this commitment at all stages from the ToR to the final report. This adherence should be meaningful and actionable, not merely procedural, with a strong focus on data protection for informants and respect for diversity.

Linked to: Conclusions 1, 2 and 6

Priority: Very high

Resource implication: Not foreseen

Timeline: Short term

Responsibility: Central Evaluation Division, evaluation managers and RPMEOs,

### **Recommendation 2**

Continue to advocate for the inclusion of gender equality and RBA in programming as cross-cutting issues while promoting environmental sustainability and disability inclusion as well, as further cross-cutting issues to be integrated into the evaluation process. This integration should begin with early needs assessments, stakeholder identification and the development

of the evaluation methodology, continuing through the findings and recommendations sections. To achieve this:

- Reinforce the support of thematic gender, human rights, environmental and disability experts throughout the evaluation process.
- Ensure the four cross-cutting issues are clearly and specifically addressed in the quality assurance tools.

Linked to: Conclusion 2

Priority: Very high

Resource implication: Not foreseen

Timeline: Short term

Responsibility: Central Evaluation Division, evaluation managers, programme managers and RPMEOs

### **Recommendation 3**

To ensure evaluations have the necessary resources to ensure high quality, adjust the evaluation budget to ensure it is proportional to the overall project budget, as recommended in the *IOM Monitoring and Evaluation Guidelines*,<sup>43</sup> as this approach appears to significantly impact ratings.

Linked to: Conclusion 5

Priority: Very high

Resource implication: Not foreseen

Timeline: Short term

Responsibility: Central Evaluation Division, project developers and PMs

### **Recommendation 4**

To enhance the utility and impact of the IOM evaluation framework, promote a culture of adaptation that enables methodological adjustments tailored to specific evaluation contexts. This approach should include:

- Designing evaluation questions that are project-specific rather than being directly extracted from previous evaluations.
- Improving sampling strategies, ensuring comprehensive triangulation, and enhancing transparency in data processes.
- Promoting a structured timeline managed through PRIMA, allowing for sufficient lead time, to enhance the robustness of results while enabling pipeline tracking and triggering alerts if calls for proposals and implementation deadlines are not met.
- Allowing flexible evaluation deliverables based on needs, such as infographics, case studies, or best practices, to improve accessibility and practical application.
- Establishing reference groups with key stakeholders for each evaluation to ensure they can effectively engage with and apply findings.
- Enhancing stakeholder engagement and fostering a learning-oriented evaluation culture to support the effective use of findings, ensuring that evaluations remain rigorous, relevant, and contextually appropriate.

---

<sup>43</sup> Two to four per cent of the overall project budget.

Linked to: Conclusions 3, 8, 10, and 11

Priority: High

Resource implication: Not foreseen

Timeline: Short term

Responsibility: Central Evaluation Division and evaluation managers

### **Recommendations to enhance terms of reference quality**

#### **Recommendation 5**

Enhance the quality of ToRs and thereby strengthen clarity, focus and methodological rigour, ultimately leading to more impactful and actionable insights. This can be achieved by:

- Including the sections “evaluation budget”, “evaluation requirements”, and “submission procedures” as mandatory in the ToR template.
- Establishing a clear procedure for the development and approval of ToRs, ensuring they are reviewed by a RPMEO or a thematic expert in alignment with quality assurance tools. This will strengthen technical oversight and enhance quality assurance.
- Standardizing the step-by-step process and distributing infographics or concise explanatory materials to help facilitate its implementation across different country offices.
- Promoting more tailored evaluation criteria, acknowledging that not all OECD-DAC criteria are always relevant and that some additional criteria might also be considered.

Linked to: Conclusions 8 and 12

Priority: Very high

Resource implication: Not foreseen

Timeline: Short term

Responsibility: Central Evaluation Division, evaluation managers and RPMEOs

### **Recommendations to strengthen the use of quality assurance tools**

#### **Recommendation 6**

Make the quality assurance tools more accessible and practical to promote their consistent use across different teams. Although these tools offer valuable guidance, their complexity and length can be challenging for some PMs, evaluation managers and evaluators, which can lead to reduced usage. To address this:

- Streamline key templates and provide more user-friendly formats, such as interactive modules or concise versions and summaries, accompanied by visuals like infographics.
- Translate these tools into languages such as French and Spanish to encourage more consistent application.
- Incorporate detailed examples of matrices that align with standards and help assess the intervention critically and objectively.
- Ensure that cross-cutting issues are addressed clearly and separately for each category (gender equality, RBA, environmental sustainability and disability inclusion) to enable focused analysis and action.
- Provide training for evaluation managers and RPMEOs on the effective use of these tools to promote their consistent application.
- Disseminate these tools among internal and external evaluators to ensure widespread adoption and adherence.

- Develop an internal mechanism to integrate quality assurance tools into workflows, embedding them into existing evaluation processes to encourage seamless adoption.
- Ensure accessibility by making quality assurance tools readily available on internal platforms, and ensure they are easily searchable and downloadable.

Linked to: Conclusions 11 and 12

Priority: Very high

Resource implication: Staff time or budget for external support

Timeline: Short term

Responsibility: Central Evaluation Division, evaluation or programme managers and RPMEOs

### **Recommendations to strengthen the capacities and skills for evaluation**

#### **Recommendation 7**

Build long-term capacity in evaluation. This requires structured and continuous training efforts; relying on one-off webinars or ad hoc learning opportunities has proven insufficient, particularly given the high staff turnover. Therefore:

- Offer step-by-step training and webinars with practical exercises and mentoring opportunities to reinforce retention and ensure methodological consistency across evaluations.
- Conduct cross-cutting issue training to ensure the genuine and consistent integration of key themes across all evaluation processes.
- Conduct training sessions and develop short tutorials to help users understand and apply quality assurance tools effectively.
- Consider decentralizing some training processes by conducting them regionally through RPMEOs or regional offices. This could foster closer engagement with PMs (future evaluation managers) and enable context-specific discussions tailored to the regional project portfolio, enhancing the practical application of evaluation principles.

Linked to: Conclusion 13

Priority: Very high

Resource implication: Not foreseen

Timeline: Short term

Responsibility: Central Evaluation Division, RPMEOs and thematic experts on RBA, gender, environment and disability.

## **LIST OF ANNEXES**

Annex I	Meta-Evaluation tool - Dimension 1 (Quality of evaluation reports)
Annex II	Meta-Evaluation tool - Dimension 2 (Comprehensiveness of evaluation reports)
Annex III	Meta-Evaluation tool - Dimension 3 (Comprehensiveness of ToR)
Annex IV	Strategies to ensure consistency
Annex V	List of documents reviewed
Annex VI	List of evaluation reports
Annex VII	ToR and inception report sampling
Annex VIII	Global scores of the components by geographic area
Annex IX	List of people interviewed
Annex X	Interview protocol
Annex XI	Questionnaire for evaluation managers
Annex XII	Questionnaire for Internal and external evaluators
Annex XIII	Terms of Reference
Annex XIV	Inception Report

## ANNEX I: Analytical framework - Dimension 1 - Quality of evaluations

IOM META-EVALUATION TOOL		
	EVALUATION INFORMATION	GUIDING COMMENTS
1.1	Report ID	
1.2	Title	
1.3	Year	
1.4	Intervention code	
1.5	Country(ies)	
1.6	Region(s)	
1.7	Sub-Region	
1.8	Thematic area	
1.9	Sub-thematic area	
1.10	Scope	Programme/Project, Strategy & Policy, Thematic
1.11	Type	Final, mid-term, etc.
1.12	Agent	Internal/External
1.13	Trained	If internal, has he/she been trained?
1.14	Geographic coverage	Country, Multi-country, Regional, Global
1.15	Author/s	
1.16	Evaluation budget	
1.18	Intervention budget	
1.19	Donor	
1.20	Evaluation Duration	Months
1.21	ToR	Availability (Yes/No)
1.22	Inception report	Yes/No - Even if only mentioned in the report
COMPONENT 1: EXECUTIVE SUMMARY – Weight 5%		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
ES.1 Explanation of the intervention background	1. Purpose 2. Objectives	Includes all the necessary information to understand the intervention and the evaluation AND does not contain information not already included in the rest of the report.
ES.2 Overview of evaluation background	1. Purpose 2. Scope 3. Intended users	
ES3. Concise description of evaluation methodology	1. Choice of methodology 2. Data sources 3. Limitations	
ES4. Summary of all evaluation findings	All evaluation findings included in the report	
ES5. Summary of all conclusions	All conclusions included in the report	
ES6. Summary of all LL & GP	All lessons learned and good practices included in the report	
ES7. Summary of all recommendations	All recommendations included in the report	
COMPONENT 2: PROJECT/PROGRAMME/EVALUAND BACKGROUND – Weight 5%		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
PB1. Brief outline of contextual factors	1. Socio-Economic context 2. Political context and institutional factors	
PB2. Clear and relevant description of key stakeholders	1. Rightholders 2. Implementing agency (ies) 3. Duty bearers/Responsibility holders	

PB3. Description of intervention/or project logic	1. Objectives and results of the intervention 2. Causal logic and assumptions 3. Locations, timelines and implementation status	The section contains relevant, accurate and complete, interconnected info reflecting the complexity of the intervention. A well-developed Theory of Change strengthens the description.
PB4. Funding arrangements	1. Reference to funding arrangements 2. Specific contributions of the IOM	
<b>COMPONENT 3: EVALUATION BACKGROUND – Weight 5%</b>		
ITEMS	ELEMENTS	RATING & COMMENTS
EB1. Explanation of purpose of evaluation	1. Purpose (s) (accountability, project improvement and/or organizational learning, etc.) 2. Explanation of evaluation purpose(s) and goals	Clear and full description of what the evaluation aims to achieve by the end of the process, including any changes made to the objectives as mentioned in the ToR (if applicable).
EB2. Description of evaluation scope	1. Geographic coverage 2. Time frame 3. Thematic coverage	If any limitations are foreseen in terms of scope, it should be mentioned here and then explained in detail in the section EM7 on limitations.
EB3. List of evaluation clients and main audiences of the report	1. Intended primary and secondary users (donors, implementing partners, etc.) 2. Intended use per client	
<b>COMPONENT 4: EVALUATION CRITERIA &amp; QUESTIONS – Weight 7,5%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
CQ1. Declaration of the evaluation criteria (e.g. OECD DAC criteria, or others) and justification for their use or lack thereof.	1. Evaluation criteria 2. Justification for their use	When using OECD DAC criteria, no need for justification as it is an international standard. Inclusion of additional or removal of criteria needs to be justified. Humanitarian evaluations should consider appropriateness, coverage; connectedness, coordination and coherence.
CQ2. Relevance of evaluation questions	1. Evaluation questions addressed the goals and purpose of the exercise 2. Questions include cross-cutting issues	At least one evaluation question should be considered for each of the criteria selected. This is clearly linked to the ToR, but the evaluator should ensure that all final evaluation questions are relevant to the purpose of the exercise. Questions about the design of the intervention should be highly considered.
CQ3. Inclusion of an evaluation matrix	1. Evaluation matrix, including evaluation indicators 2. Methods and sources of data per question included in evaluation matrix	Reference indicators and benchmarks should be included, as this is how the judgment will be justified. The evaluation matrix is usually added as an annex.
<b>COMPONENT 5: EVALUATION METHODS – Weight 7,5%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
EM1. Statement of the evaluation approach	1. Approach used 2. Justification	Approaches: Utilization Focused Evaluation, Random Control Trial, Programme Theory, Contribution analysis, Outcome mapping, participatory evaluation, etc.

EM2. Type of analysis considered	1. Type of analysis 2. Reliability assurance and levels of evidence	1. E.g. qualitative data analysis, descriptive statistics, sequential explanatory, concurrent triangulation, etc.) 2. E.g. triangulation of data collection methods, of evaluators, of sources of sampling strategies, etc.)
EM3. Description of data sources	1. List of data collection methods and explanation 2. Description of different data sources	1. E.g. document review, interviews, observation, surveys). 2. This should include a description of how diverse perspectives are captured (or if not, provide reasons for this),
EM4. Sampling procedures	1. Sampling procedures 2. Justification for the selection	The report describes the sampling frame – area and population to be represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects, and limitations of the sample.
EM5. Inclusion of cross-cutting issues	1. Gender 2. Rights-based approach 3. Disability inclusion 4. Environmental issues	The extent to which cross-cutting issues will be consider during the evaluation. The methodological strategies chosen to ensure cross-cutting issue are taken into consideration throughout the evaluation.
EM6. Stakeholder participation	1. Rationale for stakeholder participation 2. Level of participation	The evaluation report gives a complete description of stakeholder's consultation process in the evaluation. Level of participation of stakeholders is to be defined (e.g. Arnstein's ladder of participation).
EM7. Limitations of the evaluation	1. Limitations 2. Mitigation strategies	Limitations: due to the methods, due to the context, due to the sources or stakeholders, due to the evaluation team, including bias, and others. Adaptability to contextual challenges and limitations during the evaluation is something to be highly regarded.
EM8. Description of evaluation norms and standards	1. IOM & UNEG Norms & Standards 2. Ethical safeguards and data protection principles	The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards. Data protection principles explicitly included.

#### COMPONENT 6: EVALUATION FINDINGS – Weight 25%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
F1. Completeness	1. All evaluation criteria and questions are addressed 2. Findings aligned with the evaluation purpose, questions and approach	
F2. Robustness	1. Findings justified by multiple lines and levels of evidence 2. Disaggregation of data by key variables, indicators and/or	Baselines and targets are used, or their omission is justified. If the M&E framework is faulty or nonexistent,

	benchmarks as per the initiative's M&E framework.	complementary variables and indicators should be considered.
F3. Identification of causal factors leading to accomplishments and failures	1. The causal factors explaining the achievement of results or lack thereof are identified and justified 2. Description of unintended effects	The linkages of results with assumptions and other elements of the ToC strengthens the results under this component.
F4. IOM cross-cutting issues are adequately addressed	1. Gender 2. Rights-based approach 3. Disability inclusion 4. Environmental issues	
<b>COMPONENT 7: CONCLUSIONS – Weight 20%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
C1. Value	1. Conclusions are clearly linked to findings 2. Provide insights and add value to related findings	Conclusions are based on and clearly linked to the evidence presented in the findings.
C2. Reasoned	1. Conclusions reflect reasonable critical thinking and evaluative judgments 2. Judgments are to the extent possible objective	Conclusions are to the extent possible objective and clearly justified. They reflect the evaluator's own judgment regarding the evidence gathered and the reasoning process.
<b>COMPONENT 8: RECOMMENDATIONS – Weight 20%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
R1. Clarity	1. Are clear and concise 2. Are based on findings and/or conclusions of the report	1. (i.e. one or two sentences followed by explanatory text) 2. Directly and explicitly linked to findings and conclusions
R2. Relevance	1. Address key issues and are useful, tied to the object and purposes of the evaluation 2. Address cross-cutting issues	
R3. Responsibility	1. Specify who is called upon to act 2. Identify means for achievement 3. Specify priority or importance	1. Unit, department, office, etc. 2. Human, financial, material, etc. 3. Low, medium, high
R4. Actionability	1. Indicate the specific courses of action needed to remedy/or continue with the current situation 2. Provide a time frame	
<b>COMPONENT 9: LESSONS LEARNED – Weight 2.5%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
LL1. Relevance	1. Lessons learned are derived from findings or the evaluation process 2. Lessons learned represent a relevant / new piece of information to be considered in the future	The lessons deal with significant non-trivial matters. Lessons learned are generalizations based on evaluation experiences that abstract from specific circumstances to broader situations. Eventually, lessons learned can become emerging good practices.
LL2. Delimitation	1. The lessons concisely capture the context from which they were derived 2. Lessons learned target specific users	
LL3. Applicability	1. The lessons specifically suggest what should be avoided in future contexts to guide action. 2. The lessons include causal factors	

COMPONENT 10: GOOD PRACTICES – Weight 2.5%		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
GP1. Delimitation	1. The good practices concisely capture the contexts from which they were derived 2. The good practices specify target users	A good practice has been proven to work well and produce good results and is therefore recommended as a model. It is a successful experience, which has been tested and validated.
GP2. Applicability & replicability	1. The statements describe how the good practices should be implemented and by whom. 2. The good practices explore the applicability in different contexts	
GP3. Impact	1. The good practice demonstrates a link to specific impacts 2. The different impacts identified are viable / possible (realistic)	
OVERALL RATING		
An overall perception on the quality of the report to determine the extent to which it can be used with confidence.	<b>RATING (0-5) &amp; COMMENTS</b>	Describe any reason(s) that might explain the overall performance or aspects of this evaluation report. This is a chance to note mitigating factors and/or crucial issues apparent in the review of the report.

## ANNEX II: Analytical framework - Dimension 2 - Comprehensiveness of evaluation reports

DIMENSION 2 - COMPREHENSIVENESS OF EVALUATION REPORTS			
Code	ITEMS	ELEMENTS	Rating
C.1	Brief note		Yes/No
C.2	Table of Contents		Yes/No
C.3	List of tables, figures and charts		Yes/No
C.4	List of annexes		Yes/No
C.5	List of acronyms or abbreviations		Yes/No
C.6	Executive Summary		Yes/No
C.7	Intervention Background		Yes/No
C.8	Evaluation Background		Yes/No
C.9	Criteria and Questions		Yes/No
C.10	Methodology		Yes/No
C.11	Findings		Yes/No
C.12	Conclusions		Yes/No
C.13	Lessons learned		Yes/No
C.14	Emerging Good Practices		Yes/No
C.15	Recommendations		Yes/No
C.16	Annexes	A. Evaluation Matrix B. ToRs C. List of informants and site visits D. Inception report E. Data collection tools F. List of documents reviewed	Yes/No
C.17	Structure and format of the report	A. Logically structured	Yes/No

		B. Report's presentation (written in accessible way for intended audience and free from grammar, spelling and punctuation errors)	
C.18	Peer review	To be considered from 2020	Yes/No
C.19	Availability of a management response tracking tool		Yes/No

### **ANNEX III: Analytical framework - Dimension 3 - Comprehensive ness of evaluation terms of reference**

DIMENSION 3 - COMPREHENSIVENESS OF TERMS OF REFERENCE			
Code	ITEMS	ELEMENTS	Rating
ToR.1	1. Background, context	Summarizes the overall context and the project or programme that will be evaluated.	Yes/No
ToR.2	2. Description of the intervention under evaluation	The objectives and project purposes of the project or programme should be mentioned as they will be one of the main references of the evaluation.	Yes/No
ToR.3	3. Evaluation purpose	The evaluation purpose provides the underlying rationale, why the evaluation will be undertaken and how it will be used.	Yes/No
ToR.4	4. Scope and evaluation goals (Expected uses and users, what is to be done.)	The scope and objectives concretely explain what the evaluation is expected to cover and achieve. They should be clear and agreed upon by key stakeholders. The evaluation scope determines the boundaries of the evaluation, tailoring its objectives to the given situation. It should also make the coverage of the evaluation explicit (i.e. the period, phase in implementation, geographical area and the dimensions of stakeholder involvement being examined). The scope should also acknowledge the limits of the evaluation.	Yes/No
ToR.5	5. Criteria and questions	Evaluation questions, grouped by criteria, that the evaluation will need to answer.	Yes/No
ToR.6	6. Cross-cutting issues (Each one is rated separately)	- Incorporate an assessment of relevant <b>human rights, gender equality, disability inclusion and environmental issues</b> through the selection of the evaluation criteria and questions; - Specify an evaluation <b>approach and methods</b> of data collection and analysis that are rights-based and gender-responsive; - Specify that evaluation data should be <b>disaggregated</b> by social criteria (e.g. sex, ethnicity, age, disability, geographic location, income or education);	Yes/No
ToR.7	7. Methodology	Methodologies provide what information should be collected, from which source(s), for what purpose and how the collected data will be analysed to answer the evaluation questions. The methodology should include approaches, data sources, techniques, participation levels, sampling and limitations, and including the Scoring Matrix for IOMDF projects.	Yes/No
ToR.8	8. Deliverables	The ToR mention which deliverables are expected as products (reports, presentations, infographs, brief note, etc.).	Yes/No
ToR.9	9. Workplan	Detail of the estimated dates for each of the evaluation phases, and who is supposed to do what.	Yes/No
ToR.10	10. Norms and standards	Reference to norms and standards that will guide and frame the evaluation.	Yes/No
ToR.11	11. Expertise required	It could be considered an overall expertise or specific skills per consultant/team and selection of evaluators. Considers cross-cutting issues.	Yes/No
ToR.12	12. Management arrangements	Mention if there will be a reference group, who the evaluation manager/s will be, etc.	Yes/No

DIMENSION 3 - COMPREHENSIVENESS OF TERMS OF REFERENCE			
Code	ITEMS	ELEMENTS	Rating
ToR.13	13. Estimated number of days	Detail of number of days per evaluation team member including travels if they are not organized by the organization.	Yes/No
ToR.14	14. Submission procedures	Including detail of individual/company, number of people expected in the team, documents to submit, deadline.	Yes/No

## ANNEX IV: Strategies to ensure consistency

### GENERAL ISSUES

- The logic for rating items:** first, identify whether the required elements are included (even if only briefly mentioned) and then assess the level of completeness of available elements. This will help understand if the rating is within the satisfactory spectrum (elements are included) or unsatisfactory spectrum (items are missing). Then the level of completeness of existing items allow to determine the final rating. Elements included does not need to be explicitly mentioned as long as it is implicitly included in the text.
- Concerning global and thematic evaluations,** some of the items included in the tool do not necessarily entirely match (project background when we are assessing how knowledge management is conducted, for example). Nonetheless, the logic was adapted to the object of the evaluations and rated accordingly.
- Conclusions should go beyond summarizing findings and provide an interpretation of them, including an added value and the judgment. In most cases findings and conclusions might not be clearly differentiated, and usually conclusions prevail (descriptive results are not presented, but rather the whole interpretation of these results).
- In the scoring tool, “no data” was entered when information was not available and N/A when not applicable.
- ToR Comprehensiveness: cross-cutting issues:** there are three criteria for each cross-cutting issue, which is quite ambitious, include “yes” when one of them is present (most likely cross-cutting issues included in evaluation questions).
- Inception report must be mentioned in the evaluation report.

### SPECIFIC ISSUES

ITEMS	ELEMENTS	COMMENTS
ES.2 Overview of evaluation background	1. Purpose 2. Scope 3. Intended users	Implicit information: if no mention to intended users, but the purpose of the evaluation implicitly focuses on IOM as a whole or a unit. That would be a 3 (if the rest of elements are included), and it is necessary to explain that element 3 is not complete.
ES4. Summary of all evaluation findings	All evaluation findings included in the report	Consider to rate as five when a summary of all findings is included and there is a clear link between these and the findings in the report (i.e. key findings clearly highlighted in the body of the report, numbered, etc.)
ES5. Summary of all conclusions	All conclusions included in the report	Same as above. Conclusions should go beyond summarizing findings and provide an interpretation of them. Please notice in most cases findings and conclusions might not be clearly differentiated, and usually conclusions prevail (descriptive results are not presented, but rather the whole interpretation of these results).

ITEMS	ELEMENTS	COMMENTS
ES6. Summary of all LL & GP	All lessons learned and good practices included in the report	Same as ES4. Most ToR mention lessons learned and good practices, and most of reports do not include them. Include a comment when relevant in the comments section of the report.
ES7. Summary of all recommendations	All recommendations included in the report	Same as ES4. When not all recommendations are included, but reference to the full list in the report is made, rate with a 3. If recommendations are exactly the same as in the report, it can be a 5.
<b>COMPONENT 2: PROJECT BACKGROUND – Weight 5%</b>		
ITEMS	ELEMENTS	COMMENTS
PB1. Brief outline of contextual factors	1. Socio-Economic context 2. Political context	For global evaluations or non/project or programme evaluations (strategies, etc.), just consider that a good contextual background about the object under evaluation is provided.
PB2. Clear and relevant description of key stakeholders	1. Rightholder's 2. Implementing agency (ies) 3. Duty bearers/Responsibility holders	In the description of stakeholders, if beneficiaries, government and implementing partners are mentioned, give a 4.
PB3. Description of intervention logic	1. Objectives and results of the intervention 2. Causal logic and assumptions	The section contains relevant, accurate and complete, interconnected info reflecting the complexity of the project
PB4. Funding arrangements	1. Reference to funding arrangements 2. Specific contributions of the IOM	If the arrangements are adequately mentioned and presented, it should be a 4. For a 5 a more in-depth explanation of the arrangements will be required.
<b>COMPONENT 3: EVALUATION BACKGROUND – Weight 5%</b>		
ITEMS	ELEMENTS	COMMENTS
EB1. Explanation of purpose of evaluation	1. Purpose (s) (accountability, project improvement and/or organizational learning) 2. Explanation of evaluation purpose(s)	When the evaluation is justified because is part of the biannual evaluation plan, element 2 is covered.
EB2. Description of evaluation scope	1. Geographic coverage 2. Time frame 3. Thematic coverage	Consider implicit information: if it is clear that it is a final evaluation but no time frame is explicitly included, we give a 3 (it is understood that the whole life of the project). Thematic coverage: Difficult to be explicitly included, as it is taken for granted. Therefore, we could always consider it included, and if explicitly mentioned, completed.
<b>COMPONENT 4: EVALUATION CRITERIA &amp; QUESTIONS – Weight 7,5%</b>		
ITEMS	ELEMENTS	COMMENTS
CQ1. Declaration of the evaluation criteria	1. Evaluation criteria 2. Justification for their use	All OECD DAC criteria, no need for justification as it is an international standard. Inclusion of additional or removal of criteria needs to be justified. Other, such as humanitarian specific). Humanitarian evaluations should consider appropriateness, coverage; connectedness, coordination and coherence. OECD DAC criteria: if included the five main criteria, a 4. Only if the 5 criteria are complemented with other (design, coherence, coverage, specific section on cross-cutting issues etc.) provide a 5. If no mention to evaluation criteria, but evaluation criteria is implicitly included within evaluation questions, considered the criteria included, but not complete.

ITEMS	ELEMENTS	COMMENTS
CQ2. Relevance of evaluation questions	1. Evaluation questions addressed the goals and purpose of the exercise. 2. Questions include cross-cutting issues	At least one evaluation question should be considered for each of the criteria selected. This is clearly linked to the ToR, but the evaluator should ensure that all final evaluation questions are relevant to the purpose of the exercise. Questions about the design of the intervention should be highly considered. When not all cross-cutting issues are included, but one or two are, a 3 and explain. If the evaluation of programme design issues is considered, it could be a 5 (when the rest of elements are met).
CQ3. Inclusion of an evaluation matrix	1. Evaluation matrix, including evaluation indicators 2. Methods and sources of data per question included in evaluation matrix	Reference indicators and benchmarks should be included, as this is how the judgment will be justified. If that is included give a 5. If Evaluation Matrix is mentioned it will receive at least a 1. If the Evaluation Matrix has the main aspects (questions, indicators, sources, informants) give a 4.
COMPONENT 5: EVALUATION METHODS – Weight 7,5%		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
EM1. Statement of the evaluation approach	1. Approach used 2. Justification	Approaches: Utilization Focused Evaluation, Non-experimental, Random Control Trial, Programme Theory, Contribution analysis, Outcome mapping, participatory evaluation, etc. It is possible that no methods will be mentioned; must be included in comments.
EM2. Type of analysis considered	1. Type of analysis 2. Reliability assurance	1. E.g. qualitative data analysis, descriptive statistics, sequential explanatory, sequential exploratory, concurrent triangulation, etc.) 2. E.g. triangulation of data collection methods, of evaluators, of sources of sampling strategies, etc.)
EM3. Description of data sources.	1. List of data collection methods and explanation 2. Description of different data sources	1. E.g. document review, interviews, observation, surveys).  2. This should include a description of how diverse perspectives are captured (or if not, provide reasons for this) When a complete evaluation matrix is included, this should be rated 5.
EM4. Sampling procedures	1. Sampling procedures 2. Justification for the selection	The report describes the sampling frame – area and population to be represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects, and limitations of the sample. It gets a 4 when procedures and justification is provided, and 5 when it includes limitations of the sample
EM5. Inclusion of cross-cutting issues	1. Gender 2. Human Rights 3. Disability inclusion 4. Environmental issues	The extent to which cross-cutting issues will be consider during the evaluation. The methodological strategies chosen to ensure cross-cutting issue are taken into consideration throughout the evaluation. Attention: Cross-cutting issues must be included in the evaluation process and methods.
EM6. Stakeholder participation	1. Rationale for stakeholder participation 2. Level of participation	The evaluation report gives a complete description of stakeholder's consultation process in the evaluation. Level of participation of stakeholders is to be defined. Stakeholder participation: participation as "sources of information" always implicitly considered. – Always a "1" rating at least.

ITEMS	ELEMENTS	COMMENTS
EM7. Limitations of the evaluation	1. Limitations 2. Mitigation strategies	Limitations: due to the methods, due to the context, due to the sources or stakeholders, due to the evaluation team, including bias, and others. Adaptability to contextual challenges and limitations during the course of the evaluation is something to be highly regarded. Rate as 5 if limitations and mitigation strategies are included.
EM8. Description of evaluation norms and standards	1. IOM & UNEG Norms & Standards 2. Data protection principles	The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards. Data protection principles explicitly included. Considered “5” when specific actions are mentioned aimed at protecting data or to follow norms and standards.
<b>COMPONENT 6: EVALUATION FINDINGS – Weight 25%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
F2. Robustness	1. Findings justified by evidence 2. Disaggregation of data by key variables	Baselines and targets are used, or their omission is justified.
F4. IOM Cross-cutting issues are adequately addressed	1. Gender 2. Rights-based approach 3. Environmental issues	Consider a 3 when gender and RBA are very well covered, even though not environmental, and make a comment such as “Even though environmental issues are not considered, the other two cross-cutting issue are very well developed.”
<b>COMPONENT 7: CONCLUSIONS – Weight 20%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
C1. Value	1. Conclusions are clearly linked to findings 2. Provide insights and add value to related findings.	Conclusions are based and clearly linked to evidence presented in the findings.
C2. Reasoned	1. Conclusions reflect reasonable critical thinking and evaluative judgments 2. Judgments are to the extent possible objective	Conclusion are to the extent possible objective and clearly justified
<b>COMPONENT 8: RECOMMENDATIONS – Weight 20%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
R1. Clarity	1. Are clear and concise 2. Are based on findings and/or conclusions of the report	1. (i.e. one or two sentences followed by explanatory text, if needed) 2. Directly linked to findings and conclusions
R2. Relevance	1. Address key issues and are useful, tied to the object and purposes of the evaluation. 2. Address cross-cutting issues	For element 2, consider included when at least one cross-cutting issues is explicitly included in a recommendation. Explain in the comments section.
R3. Responsibility	1. Specify who is called upon to act 2. Identify means for achievement 3. Specify priority or importance	1. Unit, department, office, etc. 2. Human, financial, material. 3. Low, medium, high.
<b>COMPONENT 9: LESSONS LEARNED – Weight 2.5%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS

ITEMS	ELEMENTS	COMMENTS
LL1. Relevance	1. Lessons learned are derived from findings 2. LL represent a relevant / new piece of information to be considered in the future.	The lessons deal with significant non-trivial matters. Summarize knowledge or understanding gained from experience related to the intervention under evaluation <b>“Lessons learned</b> are generalizations based on evaluation experiences that abstract from specific circumstances to broader situations”. Please notice LL are usually required in ToR but not addressed specifically in the report. Make a comment on that when this is identified.
<b>COMPONENT 10: GOOD PRACTICES – Weight 2.5%</b>		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
GP1. Delimitation	1. The good practices concisely capture the contexts from which they were derived. 2. The good practices specify target users	“A good practice is not only a practice that is good, but a practice that has been proven to work well and produce good results, and is therefore recommended as a model. It is a successful experience, which has been tested and validated, in the broad sense, which has been repeated and deserves to be shared so that a greater number of people can adopt it”.
GP2. Applicability & replicability	1. The statements describe how the good practices should be implemented and by whom. 2. The good practices explore the applicability in different contexts.	Please notice GP are usually required in ToR but not addressed specifically in the report. Make a comment on that when this is identified.
<b>OVERALL RATING</b>		
An overall perception on the quality of the report to determine the extent to which it can be used with confidence.	RATING (0-5) & COMMENTS	Describe any reason(s) that might explain the overall performance or particular aspects of this evaluation report. This is a chance to note mitigating factors and/or crucial issues apparent in the review of the report.

#### **DIMENSION 2: COMPREHENSIVENESS OF EVALUATION REPORTS**

ITEMS	ELEMENTS	COMMENTS
9-10. Cross-cutting issues	1. Gender 2. RBA 3. Disability inclusion 4. Environment	The analysis of the cross-cutting issues will be conducted separately for each approach and will be answered with a Yes or No. This will be analysed through the inclusion of specific evaluation questions or a concrete methodology.

## ANNEX V: List of documents reviewed

YEAR	RELEVANT IOM DOCUMENTS
2006	IOM Evaluation Guidelines
2015	IOM Gender Equality Policy 2015–2019
2017	IOM Gender & RBM Guidance
2017	IOM Project Handbook second edition – Module 6 (Evaluation)
2018	IOM Evaluation Policy
2018	OIG Strategy for the Management of its Evaluation and Monitoring Functions 2018 – 2020
2018	IOM M&E Guidance: Developing an Evaluation Brief
2018	IOM Evaluation Brief Template
2018	IOM Guidance for Addressing Gender in Evaluations
2018	IOM Monitoring Policy
2019	IOM PPR Guide and annexes: Planning, Conducting and Using Project Performance Reviews
2019	MOPAN 2017-18 Performance Assessments, International Organization for Migrations
-	IOM Gender and Evaluation Tip Sheet
-	IOM Gender and RBM Tip Sheet
2020	IOM Meta-Evaluation Report 2017-2019
2020	IOM Monitoring and Evaluation Guidelines
2021	IOM Guidance for Evaluators
2021	IOM Guidance for evaluation managers
2021	OIG Monitoring and Evaluation Strategy 2021-2023
2022	Feasibility Study on Quality Management Mechanism
2022	IOM Guidance on the use of Evaluations and Follow-up of Evaluation Recommendations
2022	IOM Guidance on Quality Management of IOM Evaluations
2023	MOPAN 2021-22 Performance Assessments, International Organization for Migrations
2023	IOM Development Fund Project Evaluations
2024	IOM CED Evaluation Strategy 2024-2028
2024	IOM Evaluation Policy (draft)
RELEVANT UN DOCUMENTS	
2008	UNEG Ethical Guidelines for Evaluation
2008	UNEG Code of Conduct for Evaluation in the UN System
2010	UNEG Quality Checklist for Evaluation ToR and Inception Reports
2010	UNEG Quality Checklist for Evaluation Reports
2011	UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation
2013	UNEG Impact Evaluation in UN Agency Evaluation Systems
2014	UNEG: Integrating Human Rights and Gender Equality in Evaluation
2014	UNEG Handbook for Conducting Evaluations of Normative Work in the UN System
2016	UNEG Norms and Standards for Evaluation
2017	ILO Policy Guidelines for Evaluation
2018	UNEG SWAP Evaluation Performance Indicator Technical Note
2021	UNEG/OECD Peer Review of the International Organization for Migration (IOM) Evaluation Function
2022	UNEG Guidance on Integrating Disability Inclusion in Evaluation
2023	UNEG Guidance on Integrating Environmental Considerations in Evaluation
RELEVANT DOCUMENTS FROM OTHER ORGANIZATIONS	
2005	Assessing the Quality of Humanitarian Evaluations, The ALNAP Quality Proforma 2005

META EVALUATIONS	
2005	ALNAP Meta evaluation
2015	Meta evaluation on Finland's Development Policy and Co-operation 2012-2014
2017	GEROS <sup>44</sup> Handbook for UNICEF Staff & Independent Assessors
2018	UNICEF Geros Meta-Analysis 2017
2018	UN Women Meta evaluation 2017 (GERAAS)
2019	GEROS Evaluation Quality Assurance Tool
2020	Meta-Evaluation of IOM's internal and external evaluations (2017 – 2019)
2024	Meta-evaluation of IOM Development Fund projects (funded 2016-2021)

<sup>44</sup> "GERO" refers to UNICEF's Global Evaluation Reports Oversight System, which manages quality assurance and public accessibility of the organization's evaluation reports.

## ANNEX VI: List of evaluations

Evaluation reports assessed with the meta-evaluation tools - Dimension 1 - Quality of evaluations, and - Dimension 2 - Comprehensiveness of evaluation reports.

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
1	2020		Ex-Post Evaluation Report "Enhancing the Development Impact of Engagement with The Moldovan Diaspora – Homeland Partnerships"	IOM Development Fund	Country	Republic of Moldova	Europe	Programme/Project	Internal	Migration and Development
2	2020	DP.1291	Programme evaluation report	Government of the United Kingdom	Country	Pakistan	Asia & Pacific	Programme/Project	External	Preparedness and Response
3	2020	TC.1052	Ex-post internal evaluation report for IDF project: "Strengthening the capacity of the government of Kenya to manage national identity programmes"	IOM Development Fund	Country	Kenya	Africa	Programme/Project	Internal	Border and identity solutions
4	2020	LM.0309	Ex-post evaluation: technical support to the bureau of manpower	IOM Development Fund	Country	Bangladesh	Asia & Pacific	Programme/Project	External	Labour migration
5	2020	LM.0279	Ex-post evaluation of the IOM project "Supporting strategic interventions towards effective and sustainable labour migration management and information sharing in Tanzania"	IOM Development Fund	Country	United Republic of Tanzania	Africa	Programme/Project	External	Labour migration
6	2020	IB.0122	Final evaluation of the project "Improving information management and planning capacities of Serbian commissariat for refugees and migration"	Government of Switzerland	Country	Serbia	Europe	Programme/Project	External	Border and identity solutions

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
7	2020	RT.1372 RT.1373 RT.1376	Final evaluation of the project "Consolidate direct assistance and assisted voluntary return and reintegration services in Egypt (care)"	Government of Germany	Country	Comoros, Egypt, Ethiopia, Sudan	Africa	Programme/Project	External	Assistance to vulnerable migrants, AVRR
8	2020	MA.0379	Ex-post Evaluation: Enhancing Mechanisms for Prevention	IOM Development Fund	Multi-country	Armenia, Azerbaijan, Georgia	Europe	Programme/Project	External	Migration health
9	2020	CS1028 CS0968	Final internal evaluation of the "Tajik Afghan integration"	Government of the United Kingdom	Country	Afghanistan, Tajikistan	Asia & Pacific	Programme/Project	Internal	Multiple thematic overview
10	2020	CT1043	Final evaluation of the Project "improving knowledge enforcement and coordination in counter-trafficking"	Government of the United States	Country	Rwanda	Africa	Programme/Project	External	Counter trafficking
11	2020		Evaluation of the World Migration Report	IOM	Global	Worldwide	Worldwide	Policy / Thematic	Central /OIG	Multiple thematic overview
12	2020	CE.0316	Final evaluation of the project "Migration as a livelihood diversification strategy in the delta (MILDAS)"	UNOPS	Country	Myanmar	Asia & Pacific	Programme/Project	External	Migration and Development
13	2020	MA.0436	Final evaluation for the project: "Psychosocial and medical assistance to urban refugee survivors and those at risk of sexual and gender-based violence (SGBV) in Kampala"	Government of the United States	Country	Uganda	Africa	Programme/Project	Internal	Migration health, Media and communications
14	2020	CT.1145	"Strengthening the identification and protection of victims of trafficking among refugees in Rwanda"	Government of the United States	Country	Rwanda	Africa	Programme/Project	External	Counter trafficking
15	2020	TC.0829	Ex-Post Evaluation of the Project "ICT Strategy for Integrated Border Management in Rwanda"	IOM Development Fund	Country	Rwanda	Africa	Programme/Project	Internal	Border and Identify solutions

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
16	2020	LM. 0325	Strengthening labour migration management in Botswana	IOM Development Fund	Country	Botswana	Africa	Programme/Project	Internal	Labour migration
17	2020	CT.1007	Évaluation Finale Externe du Projet : Renforcement des Capacités pour une Meilleure Coordination	Government of the United States	Country	Madagascar	Africa	Programme/Project	External	Counter trafficking
18	2020	CT.0914	Dignity and rights in Central Asia final performance evaluation	Government of the United States	Country	Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan	Asia & Pacific	Programme/Project	External	Multiple thematic overview
19	2020	DC.0005	Enhancing Human Security and Building Resilient Societies in Disadvantaged Communities of Armenia	UN Agencies	Country	Armenia	Asia & Pacific	Programme/Project	External	Migration and Development
20	2020	LM.0340	Safeguarding the workers on the move in the MENA region	IOM	Multi-country	Egypt, Jordan, Morocco	Africa	Programme/Project	Internal	Labour migration
21	2020	CT.1168	Ex-post evaluation of the project 'strengthening the national capacities in counter trafficking in human beings in North Macedonia'	IOM Development Fund	Country	North Macedonia	Europe	Programme/Project	External	Counter trafficking
22	2020	CE.0376	External Evaluation of the Project 'Institutionalize Health Care Improvement through Temporary Returns of Somali Diaspora Health Professionals to Somaliland through Migration for Development in Africa (MIDA)	Government of Finland	Country	Somalia	Africa	Programme/Project	External	Migration and Development
23	2020	CS. 0948	Évaluation finale du projet "Initiatives de Stabilisation Communautaire dans le Nord du Niger (COSINN)	Government of Germany	Country	Niger	Africa	Programme/Project	External	Transition Recovery

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
24	2020	PR.0239	Internal independent final evaluation of programme migration for development II - supporting the Albanian institute of statistics (INSTAT) in conducting a national household migration survey (HMS)	Government of Germany	Country	Albania	Europe	Programme/Project	Internal	Migration research
25	2020	CS.0607	Wave III: Community Perception Survey in Diffa	Government of the United States	Country	Niger	Africa	Programme/Project	External	Transition Recovery
26	2020		Second evaluation of IOM's migration initiatives	Not available	Global	Worldwide	Worldwide	Strategy	Central /OIG	Multiple thematic overview
27	2020	CE.0355	Project: external review of «mainstreaming the concept of migration and development into strategies	Government of Switzerland	Country	Bosnia and Herzegovina	Europe	Programme/Project	External	Migration and Development
28	2020	CS.0936	Ex-post internal evaluation of the project "community-based approach to support youth in a targeted municipality"	IOM Development Fund	Country	North Macedonia	Europe	Programme/Project	Internal	Transition Recovery
29	2020	RT.1297	Fortalecimiento del proceso de integración de migrantes que retornan a El Salvador (PIRE)	IOM Development Fund	Country	El Salvador	America & the Caribbean	Programme/Project	External	AVRR
30	2020	DS.0017	Mid-Term Evaluation of MIDNIMO 2 Project: Support for the Attainment of Durable Solutions in Areas Impacted by Displacement and Returns	UN Peace Fund	Country	Somalia	Africa	Programme/Project	External	Transition Recovery
31	2020	IM.0043	Ex-post evaluation of the IOM development fund project	IOM Development Fund	Country	Brazil	America & the Caribbean	Programme/Project	External	Labour migration
32	2020		Final evaluation of MIDA Somalia phase I	Government of Italy	Country	Somalia	Africa	Programme/Project	External	Migration and Development
33	2021	CT.1139	Solomon Islands: protecting the rights of women and children	European Union	Country	Solomon Island	Asia & Pacific	Programme/Project	External	Counter trafficking

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
34	2021	MA.1147	Ex-post Evaluation: Strengthening Government Capacity in the Development and Implementation of the National Strategic Action Plan on Migration Health in Nepal	IOM Development Fund	Country	Nepal	Asia & Pacific	Programme/Project	External	Migration health
35	2021	DP.2083	Final Evaluation: "strengthening mechanisms for humanitarian assistance to Venezuelan migrants	Government of Japan	Country	Colombia	America & the Caribbean	Programme/Project	External	
36	2021	PB.0043	Evaluación final de la estrategia de fortalecimiento institucional para el despliegue y funcionamiento territorial articulado del sistema integral de verdad	Government of Colombia	Country	Colombia	America & the Caribbean	Programme/Project	External	Transition Recovery
37	2021	CS.1079	Evaluación de resultados del proyecto de "fortalecimiento de la estrategia de protección colectiva & prevención de la violencia rural	Government of the United Kingdom	Country	Colombia	America & the Caribbean	Programme/Project	External	Transition Recovery
38	2021	RP.0110	Evaluación final externa del proyecto "contribuir al mejoramiento de la atención de víctimas y de gestión de conocimiento en los procesos de administración de justicia a través del desarrollo de capacidades y asistencia técnica a la justicia	Government of Colombia	Country	Colombia	America & the Caribbean	Programme/Project	External	Transition Recovery
39	2021	DR.0008	Informe de Evaluación Final USAID Tropical Storm IDA Reconstruction Project JMPL Schools	Government of the United States	Country	El Salvador	America & the Caribbean	Programme/Project	External	Transition Recovery
40	2021	MP.0409	Final evaluation of expanding the protection and promotion of mental health and psychosocial support for communities affected by the refugee emergency in Cox's Bazar	Government of Sweden	Country	Bangladesh	Asia & Pacific	Programme/Project	External	Migration health
41	2021	LM.0327	Final Evaluation: Capacity Building for Sustainable Development in Tonga	IOM Development Fund	Country	Kingdom of Tonga	Asia & Pacific	Programme/Project	External	Migration and climate change, Migration policy

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
42	2021	PB.0038	Strengthening of territorial capacities for monitoring the Implementation Framework Plan and managing compliance with the commitments established in the Final Peace Agreement - SI IPO	Government of Colombia	Country	Colombia	America & the Caribbean	Programme/Project	External	Transition Recovery
43	2021	PB.0011	Programa para el fortalecimiento del sistema integral de verdad	Government of Sweden	Country	Colombia	America & the Caribbean	Programme/Project	External	Transition Recovery
44	2021	IS.0051	Evaluación final del proyecto "apoyo a la integración socioeconómica y fortalecimiento de los sistemas de salud y de las organizaciones de la sociedad civil en el contexto de COVID-19 y su impacto en la población venezolana y receptora" en Ecuador	Government of Spain	Country	Ecuador	America & the Caribbean	Programme/Project	External	Migration and Development
45	2021	DP.2122 DP.2110	Final external evaluation of "promoting rights-based solutions for vulnerable migrants through a migrant resource and response mechanism" (MRRM) programmes"	Government the Netherlands, Government Switzerland	Country	Libya	Africa	Programme/Project	External	Assistance to vulnerable migrants
46	2021	CO10P052 7 MA.0462	Evaluación de resultados del proyecto salud para la paz	The Multi-Partner Trust Fund Office (MPTF)	Country	Colombia	America & the Caribbean	Programme/Project	External	Transition Recovery
47	2021	MX10P05 03 PX.0087	Evaluación Final del proyecto "Fortalecimiento de las capacidades del gobierno estatal"	Not available	Country	Mexico	America & the Caribbean	Programme/Project	External	Counter trafficking
48	2021	LM.0419	External Mid-Term Evaluation of the Western Hemisphere Program	Government of the United States	Regional	Multi country <sup>45</sup>	America & the Caribbean	Programme/Project	External	Multiple thematic overview

<sup>45</sup> Belize, Costa Rica, Dominica, Dominican Republic, El Salvador, Guatemala, Guyana, Haiti, Honduras, Mexico, Panama, The Bahamas.

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
49	2021	RR.0043	Nigeria: strengthening assistance for returnees and potential migrants and promoting safe migration practices in communities of origin	Government of Germany	Country	Nigeria	Africa	Programme/Project	Internal	Assistance to vulnerable migrants, AVRR
50	2021	DP.2218	Evaluación intermedia e identificación de lecciones y buenas prácticas del proyecto “fortalecimiento de la respuesta regional a la migración a gran escala de los nacionales venezolanos en Ecuador”	Government of the United States	Country	Ecuador	America & the Caribbean	Programme/Project	External	Preparedness and Response
51	2021	CT.1150	Internal final evaluation of the project: “enhancing capacities to fight trafficking in persons in Niger”	Government of the United States	Country	Niger	Africa	Programme/Project	Internal	Assistance to vulnerable migrants, Counter trafficking
52	2021		Regional Development and Protection Programme for North Africa (RDPP NA) A Success Case Method (SCM) Evaluation Study	European Union, Government Austria, Government Italy, Government Norway, Government the Czech Republic	Regional	Algeria, Chad, Egypt, Italy, Liberia, Mauritania, Morocco, Niger, Tunisia	Africa	Programme/Project	External	Multiple thematic overview
53	2021	CS.0815	Strengthening social cohesion and stability in slum populations	European Union	Country	Uganda	Africa	Programme/Project	External	Transition Recovery
54	2021	IS.0011	Final external independent evaluation of commit project: facilitating the integration of resettled refugees in Croatia, Italy Spain and Portugal	European Union	Multi-country	Croatia, Italy, Portugal, Spain	Europe	Programme/Project	External	Multiple thematic overview

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
55	2021	RE.0356	Evaluación final del Proyecto de fortalecimiento de la protección internacional recepción e integración de refugiados en Argentina	European Union	Country	Argentina	America & the Caribbean	Programme/ Project	External	Resettlement and Movement Management (RMM)
56	2021	PB.0021	Strengthening the role of Women and Youth as Peace Builders to improve Development in the Highlands of Papua New Guinea	UN Peace Fund	Country	Papua New Guinea	Asia & Pacific	Programme/ Project	External	Transition Recovery
57	2021	CE.0350	Ex-post Evaluation: Strengthening Internal Migration Policies in Bolivia	IOM Development Fund	Country	Ecuador	America & the Caribbean	Programme/ Project	External	Assistance to vulnerable migrants, Migration and Development
58	2021	PR.0194	Ex-post Evaluation : Migration Profile - Rwanda	IOM Development Fund	Country	Rwanda	Africa	Programme/ Project	External	Migration research
59	2021	IS.0051	Evaluación final del proyecto "apoyo a la integración socioeconómica y fortalecimiento de los sistemas de salud y de las organizaciones de la sociedad civil en el contexto de COVID-19 y su impacto en la población venezolana y receptora" en Perú	Government of Spain	Country	Peru	America & the Caribbean	Programme/ Project	External	Migration and Development
60	2021		Evaluation of IOM's Institutional Response to Address Migration, Environment and Climate Change Nexus	Not specified	Global	Worldwide	Worldwide	Thematic	Central /IOG	Migration and climate change
61	2022	MM.0040	External Evaluation of IOM's Approach to Protection Against Sexual Exploitation and Abuse and Sexual Harassment	IOM	Global	Worldwide	Worldwide	Thematic	External	Organisational policy/strategy

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
62	2022	IB.0080	Ex-Post Evaluation of the Migrant Registration Support for Tanzania (MIREG) project	Government of the United Kingdom	Country	United Republic of Tanzania	Africa	Programme/Project	Internal	Border and identity solutions
63	2022	CS.1122	Socio-Economic Reintegration of Returnees and other Vulnerable Households in migration affected areas severely impacted by COVID-19 Pandemic	Government of Japan	Country	Lesotho	Africa	Programme/Project	External	AVRR
64	2022	TC.0994	Informe de la evaluación final para el Proyecto de “Fortalecimiento de capacidades para prevenir y combatir el tráfico ilícito de migrantes por mar en México”.	Government of Canada	Country	Mexico	America & the Caribbean	Programme/Project	External	Counter trafficking
65	2022	IV.0011	Mid-term evaluation of “Australia proposing organisation for community support programmes	Not available	Country	Australia	Asia & Pacific	Programme/Project	Internal	Assistance to vulnerable migrants
66	2022	DP.2458	Evaluación temática de la asistencia humanitaria de emergencia de la OIM a personas refugiadas y migrantes de República Bolivariana de Venezuela en Ecuador	Government of the United States	Country	Ecuador	America & the Caribbean	Programme/Project	External	Preparedness and Response
67	2022	CS.1075	Evaluation final du projet « comprendre et traiter les facteurs de conflits le long des routes migratoires d'Agadez »	UN Peace Fund	Country	Niger	Africa	Programme/Project	External	Transition Recovery
68	2022	CD.0035	Evaluation ex-post du projet « Soutenir la jeunesse avec des centres de conseil en emploi efficents et efficaces au Niger »	IOM Development Fund	Country	Niger	Africa	Programme/Project	External	Migration and Development
69	2022	HN10P00 01 CT.0938	External ex-post Evaluation: Building the Capacity of the Honduran Consular Service for the Protection of Honduran Migrants Transiting through Mexico	IOM Development Fund	Country	Honduras	America & the Caribbean	Programme/Project	External	Assistance to vulnerable migrants, Migration and Development

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
70	2022	PX.0121	Evaluación externa final proyecto: fortalecimiento de las organizaciones de sociedad civil para la prevención e identificación de posibles casos de trata de personas a nivel comunitario	Government of Switzerland	Country	Nicaragua	America & the Caribbean	Programme/ Project	Externally	Assistance to vulnerable migrants, Counter trafficking
71	2022	RQ.0029	MIDA FINNSOM Health and Education Project Phase II	Government of Finland	Country	Somalia	Africa	Programme/ Project	Externally	Migration and Development
72	2022	RR.0116	Evaluación del proyecto UNJP/GUA/035/PBF Construir la cohesión social de las comunidades que reciben jóvenes retornados como un puente hacia una reintegración pacífica y efectiva	UN Peace Fund	Country	Guatemala	America & the Caribbean	Programme/ Project	Externally	Transition Recovery
73	2022	IS.0051	Evaluación final del proyecto "apoyo a la integración socioeconómica y fortalecimiento de los sistemas de salud y de las organizaciones de la sociedad civil en el contexto de COVID-19 y su impacto en la población venezolana y receptora"	Government of Spain	Country	Venezuela (Bolivarian Republic of)	America & the Caribbean	Programme/ Project	Externally	Migration and Development
74	2022	TC.1028	Evaluation finale de l'initiative conjointe UE-IOM pour le renforcement de la gestion des frontières	European Union	Country	Mauritania	Africa	Programme/ Project	Externally	Assistance to vulnerable migrants
75	2022	MA.0485	Evaluation of the Fostering Health and Protection to vulnerable migrants transiting through Morocco	Government of Finland	Regional	Libya, Morocco, Sudan, Tunisia, Yemen	Africa	Programme/ Project	Externally	Assistance to vulnerable migrants
76	2022	FC.0164	Implementación de planes de prevención de violencia basada en género y promoción de la autonomía de mujeres de las FARC en tránsito a la vida civil (Fase II)	Government of Sweden	Country	Colombia	America & the Caribbean	Programme/ Project	Externally	Transition Recovery

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
77	2022		Final evaluation of the EU-IOM knowledge management hub (KMH)	European Union	Global	Worldwide	Worldwide	Programme/Project	External	AVRR
78	2022	IB.0133	External ex-post combined Evaluation: of IOM project support for the implementation of the Migration Information and Data Analysis System (MIDAS)	Government of Paraguay	Country	Paraguay	America & the Caribbean	Programme/Project	External	Border and identity solutions
79	2022	PX.0100	Informe de Evaluación Intermedia del Proyecto Pionero de Inserción Social	Government of the United States	Country	El Salvador	America & the Caribbean	Programme/Project	External	Assistance to vulnerable migrants
80	2022	PM.0021	Evaluación final del proyecto “promoción de contribuciones positivas de la migración en América del sur a través de los medios de comunicación y su interacción con OIM”	IOM	Country	Argentina	America & the Caribbean	Programme/Project	External	Media and communications
81	2023	DR.0040	Final Internal Project Evaluation Reduction of Economic and Human Loss and Displacement by Natural Disasters through Community-based Resilience building and the Capacity-building of the Government at All Levels in Papua New Guinea	Government of the United States	Country	Papua New Guinea	Asia & Pacific	Programme/Project	Internal	Preparedness and Response
82	2023	DP.2377	Final External Evaluation of “Protecting vulnerable migrants and stabilizing communities in Libya – Phase II” (EUTF) Program	European Union	Country	Libya	Africa	Programme/Project	External	Multiple thematic overview
83	2023	CS.1156	Final evaluation of the project: “Bosnia and Herzegovina: Enhancing social cohesion in communities hosting people on the move”		Country	Bosnia and Herzegovina	Europe	Programme/Project	External	Transition Recovery

No	Year	Code	Title	Donor	Coverage	Countries	Region	Scope	Agent	Thematic Areas
84	2023	RT.1354	Final Evaluation of the EU-IOM Joint Initiative for migrant protection and reintegration in the horn of Africa	European Union	Regional	Djibouti, Ethiopia, Somalia, South Sudan	Africa	Programme/ Project	External	Assistance to vulnerable migrants, AVRR
85	2023	RR.0075	Informe Inicial: Evaluación Ex Post del Proyecto de Fortalecimiento de los procesos de retorno y reintegración en el Paraguay	IOM Development Fund	Country	Paraguay	America & the Caribbean	Programme/ Project	External	AVRR
86	2023		Thematic evaluation of IOM's labour migration and mobility strategy and initiatives	IOM	Global	Worldwide	Worldwide	Thematic	Central /OIG	Labour migration
87	2023	T085-EUTF-NOA-LY-03-01 T05.141	Final evaluation of the EU-IOM Joint Initiative for Migrant Protection and Reintegration-North Africa	European Union	Multi-country	Algeria, Egypt, Libya, Morocco, Tunisia	Africa	Programme/ Project	External	Assistance to vulnerable migrants, AVRR
88	2024	DP.0970	Mid-term evaluation of the project: Addressing Mixed Migration Flows in the Western Balkans	Government of Norway	Regional	Albania, Bosnia and Herzegovina Montenegro North Macedonia, Serbia, UNSC Resolution - administered Kosovo	Europe	Programme/ Project	External	Multiple thematic overview
89	2024		Evaluation of IOM's contribution to the migration dialogue for Southern Africa (MIDSA)	IOM	Global	Worldwide	Worldwide	Policy	Central /OIG	Multiple thematic overview
90	2024	NC.0073	Increasing the Knowledge Base on Community Cohesion and Mobility Dynamics in the Context of Climate Change and Environmental Degradation through Select Country-based Case Studies within the MENA Region	IOM Development Fund	Country	Libya, Sudan	Africa	Programme/ Project	Internal	Multiple thematic overview

## **ANNEX VII: List of terms of reference and inception report sampling**

ToRs assessed against the IOM checklist for terms of reference, including all ToRs attached to the sampled evaluation reports:

N°	Year	N°	Year
1	2020	35	2021
2	2020	36	2021
3	2020	40	2021
4	2020	41	2021
5	2020	43	2021
6	2020	44	2021
7	2020	46	2021
8	2020	50	2021
9	2020	52	2021
10	2020	55	2021
11	2020	56	2021
12	2020	58	2021
14	2020	59	2021
15	2020	63	2022
16	2020	64	2022
17	2020	65	2022
18	2020	68	2022
19	2020	69	2022
20	2020	75	2022
21	2020	76	2021
22	2020	77	2022
24	2020	78	2022
26	2020	79	2022
28	2020	80	2022
29	2020	81	2023
30	2020	82	2023
31	2020	85	2023
33	2021	86	2023
34	2021	89	2024

ToRs assessed with the tool -Dimension 3- Comprehensiveness of evaluation terms of reference

Nº	Year	ToR - Report Number	Global Scoring <sup>46</sup>
1	2020	18	4
2	2020	9	3,9
3	2020	3	2,7
4	2020	4	*2,1
5	2020	2	1,4
6	2021	36	4,5
7	2021	35	4,4
8	2021	50	3,1
9	2021	40	*2
10	2021	46	*1,9
11	2022	78	3,8
12	2022	69	3,5
13	2022	63	1,9
14	2022	79	*1,8
15	2023	85	*3,3
16	2023	86	3,2
17	2023	82	*2,6
18	2023	81	1,8
19	2024	89	2,4

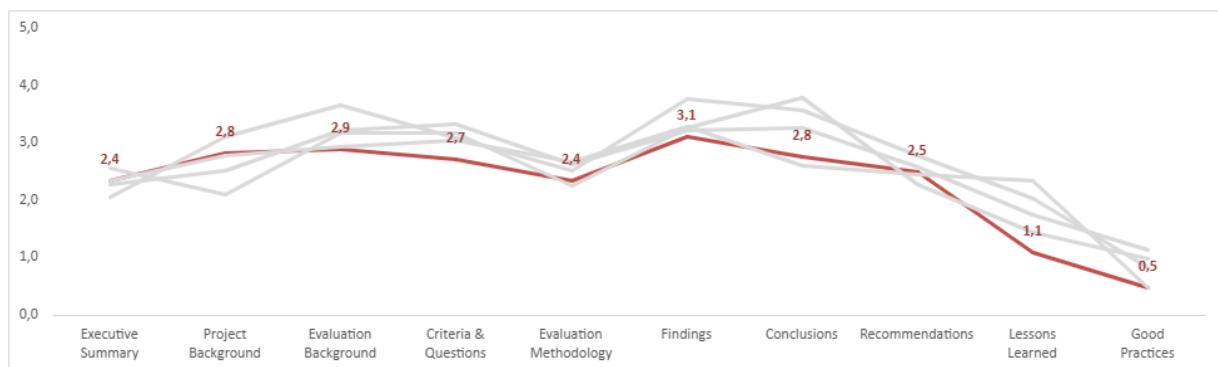
Inception reports assessed with the IOM quality control tool for inception reports

Nº	Year	Report Number
1	2020	9
2	2022	78
3	2023	86

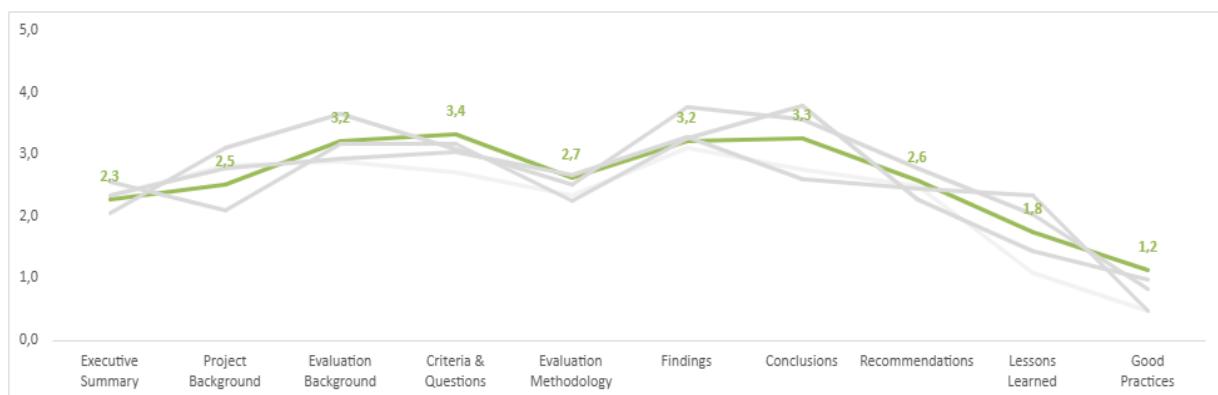
<sup>46</sup> The asterisk indicates other reports with a higher or lower score, depending on the case, which have been discarded due to the lack of ToR.

## ANNEX VII: Global scores of the components by geographic area

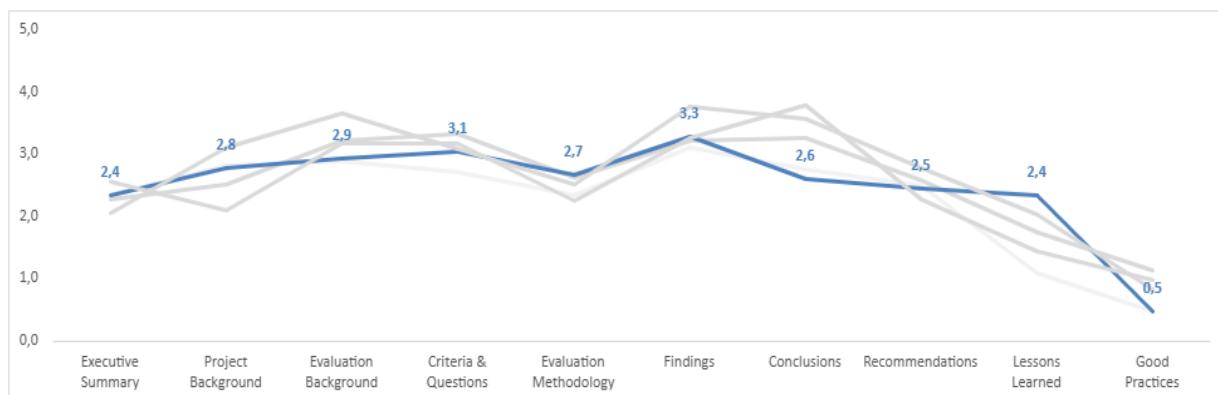
### Africa



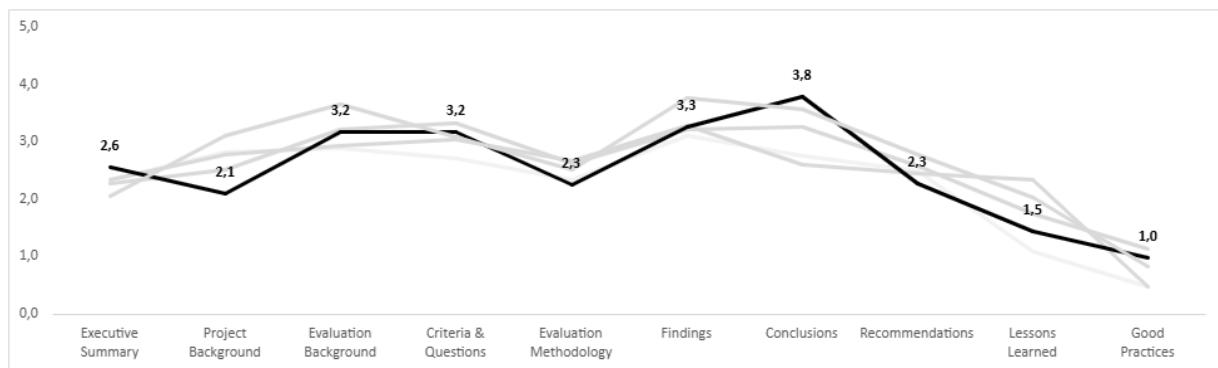
### America & the Caribbean



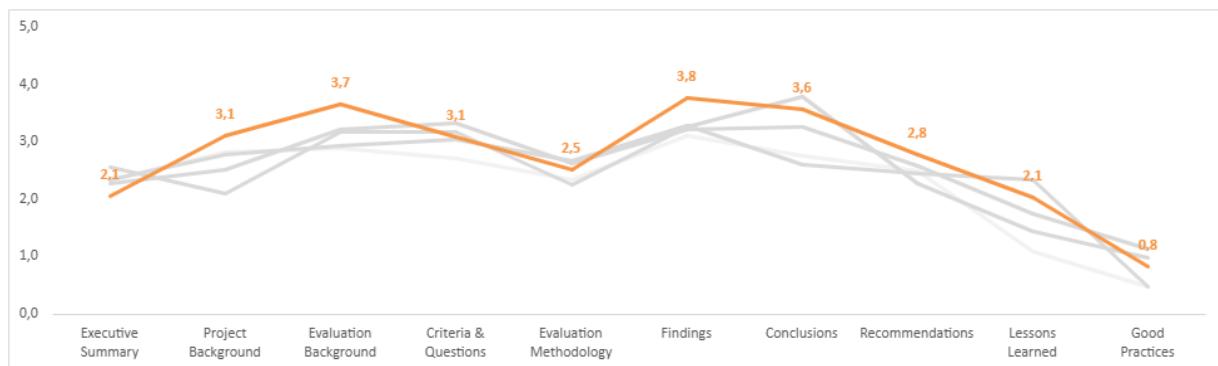
### Asia & the Pacific



### Europe



## Global



## ANNEX IX: List of people interviewed

No	Name	Role
1	Alessia CASTELFRANCO	IOM Development Fund Administrator
2	Mariana SATO*	IOM Development Fund Programme Assistant
3	Elma BALIC	Chief of Evaluation OIG
4	Muti ROGERS*	Regional M&E Officer
5	Soutongnooma Yves KOALA*	Regional M&E Officer
6	Nisreen ABOUR MRAD	Regional M&E Officer
7	Angeline WABANDA*	Regional M&E Officer
8	Theogene NSHIMIYIMANA*	Regional M&E Officer
9	Sarah HARRIS	RBM Officer
10	Zahoor AHMAD	M&E Officer
11	Diana CARTIER	Chief of Mission Zimbabwe
12	Kathia BARRECH	Oversight Officer USRAP
13	Davide BRUSCOLI	Regional Information Management Officer
14	Glen O'neill	External evaluator
15	Sebastián León Giraldo	External evaluator
16	Sarah Pugh	External evaluator

Note: Members of the evaluation reference group are indicated with an asterisk.

## ANNEX X: Interview protocol<sup>47</sup>

**1. SECTION: Preparing the evaluation** (considers questions 1, 2 and 3 of the meta-evaluation)

**AREA:** Decision-making concerning ToR/ Evaluation design/ External teams

**BROAD QUESTIONS:**

- a) What are the key processes and logic followed when preparing evaluations?
- b) What is the involvement of Headquarters (Central Evaluation Division for Headquarters) managed or global evaluations or from Country representatives for regional and country evaluations in the evaluation design? How could that involvement be strengthened/improved?
- a) To what extent are the new quality (QA) tools designed by the CED disseminated and used? - the Checklist for Terms of Reference, the Quality control tool for inception reports, the Quality control tool for evaluation reports (2022), and to the UNEG quality checklist for TORs/Inception reports and the UNEG checklist for evaluation reports (2010)- Assessment of its quality.

**INFORMANTS:** National and regional monitoring and evaluation officers

**GUIDING QUESTIONS:**

- What is the basis (why and how) to conduct an evaluation?
- How are evaluation time frames and timing decided?
- What is the basis for conducting internal or external evaluations?
- Who prepares the ToR? Would you say they are properly done (with knowledge and following the ToR-related guidelines)?
- What is your level of involvement/participation in elaborating ToR?

---

<sup>47</sup> Informants: national and regional monitoring and evaluation officers, and key informants suggested by the CED (internal and external evaluators).

- Do you know the different tools – Guidance on Quality of IOM Evaluation 2022- that have been developed to guarantee the quality of the evaluations? Have you used them? If not, why?
- How do you value the usefulness of quality tools? Why?

*INFORMANTS:* internal and external evaluators

**GUIDING QUESTIONS:**

- How are evaluation time frames and timing decided?
- Would you say ToRs are properly done (with knowledge and following the ToR-related guidelines)?
- Do you know the different tools that have been developed to guarantee the quality of the evaluations? Have you used them? If not, why?
- How do you value the usefulness of the 2022 Guidance on the quality of IOM evaluation? Why?

## 2. SECTION: During the evaluation (consider questions 2, 3, 4 and 5)

**AREA:** Participation/ Process/ Supervision

**BROAD QUESTION:**

- a) What is your assessment of the evaluation process in IOM and your specific role and contribution to it?

*INFORMANTS:* National and regional monitoring and evaluation officers

**GUIDING QUESTIONS:**

- Were you consulted during the evaluation process?
- Is there adequate supervision during the evaluation processes?
- Who selects the key informants and the documents to review?
- What happens if any difficulties arise?
- What are the processes that work best and why?
- What could be done to improve the evaluation process?
- What are the main limitations to conducting evaluations (resources, skills and competencies, time allocated, etc.)
- To what extent did ongoing efforts in M&E impact evaluation products?

*INFORMANTS:* internal and external evaluators

**GUIDING QUESTIONS:**

- What positive aspects would you highlight in the evaluation process? Is there any room for improvement?
- How would you rate the collaboration with IOM evaluation managers?
- What mechanisms do you develop to ensure that the evaluation meets IOM evaluation standards?
- How do you address cross-cutting issues? (Gender, human rights, disability, and environmental integration)
- During the evaluation, has the organization provided everything necessary to carry out a good evaluation?
- What happens if any difficulties arise?

## 3. SECTION: Level of improvement in the quality of evaluations and main trends of that quality. (Considers questions 4 and 6)

**AREA:** Analysis of the quality of evaluation reports

**BROAD QUESTION:**

- a) Is the quality of evaluations improving in IOM? (Quality assessment)

*INFORMANTS:* National and regional monitoring and evaluation officers

**GUIDING QUESTIONS:**

- Has the quality of evaluations improved in recent years? What are the key contributing factors?
- Which sections of the reports tend to be of lower quality? Why?
- Do the evaluations provide appropriate and realistic recommendations?
- What could be done to improve the use/uptake of evaluations? What is the perceived credibility of evaluations?

*INFORMANTS:* Internal and external evaluators

#### GUIDING QUESTIONS

- Which sections of your evaluation report(s) do you consider to be of the highest quality?
- What part(s) of the evaluation has been most challenging?
- Would you consider that the quality control tools provided by the IOM facilitate the assurance of the quality of reports??
- Do you think your evaluation(s) provide(s) appropriate and realistic recommendations for IOM?
- Do you consider that the IOM evaluation managers have collaborated adequately in the evaluation process?

## ANNEX XI: Survey protocol - Evaluation managers

### DEMOGRAPHICS & CONTROL QUESTIONS

#### 1. Gender

- Female  
 Male  
 Other gender identities

#### 2. Years working at IOM

- Less than 1 year  
 1 to 3 years  
 4 to 6 years  
 7 to 9 years  
 10 and more years

#### 3. Category of your current position in IOM

"General"/G Staff

"National Officer"/NO Staff

"Professional"/P staff

"Director"/D staff

Consultant

Intern

Other: \_\_\_\_\_

#### 4. Location of post:

- North, Central and South America and the Caribbean (RO Panama)  
 Europa and Central Asia (RO Vienna)  
 East, Horn and Southern Africa (RO Nairobi)  
 North Africa and Middle East (RO Cairo)  
 West and Central Africa (RO Dakar)  
 Asia and the Pacific (RO Bangkok)  
 Global Office - Washington  
 Global Office - Brussels  
 Headquarters (Geneva)

#### 5. The Evaluation Manager is the person who designs the evaluation, selects the evaluator, oversees the conduct of the evaluation, and reviews the evaluation deliverables. Please indicate the number of evaluations where you have been the Evaluation Manager in IOM since 2020 for each of the following situations (Please mark all that apply)

Evaluation manager with no formal role or involvement in the evaluated project

Evaluation manager while also serving as PM of the evaluated project

Evaluation manager while also supervising the PM of the evaluated project.

Evaluation manager while also being a team member of the evaluated project

Evaluation manager while part of the same office implementing the project, but without direct project involvement

Evaluation manager with no formal role or involvement in the evaluated project

I have not been evaluator manager during the designated period (-> then end of the survey)

Other, please specify:

## SECTION 1 – EVALUATION PROCESS AND INVOLVEMENT

1. To what extent do you think the quality of evaluation ToR has improved in the last 5 years?

Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	N/A
<input type="checkbox"/>						

2. Are the IOM evaluation guidelines, reference materials, and methodological guidelines (Quality Control tools) sufficient and clear enough to:

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	N/A
Draft, review and adjust the Evaluation ToR	<input type="checkbox"/>						
Guide, control and adjust Inception Reports	<input type="checkbox"/>						
Manage and ensure quality of evaluations	<input type="checkbox"/>						

3. To what extent do you consider the process of reviewing and discussing the report(s) to be sufficiently inclusive and participatory to gather the most relevant information for the analysis?

- Strongly agree
- Agree
- Somewhat agree
- Somewhat disagree
- Disagree
- Strongly disagree
- Usually, no comments are received

4. For the evaluations you managed since 2020, how consistently were the following cross-cutting issues integrated into the evaluation process? Mark all that apply.

	Always	Usually	Sometimes	Rarely	Never	N/A
Gender mainstreaming	<input type="checkbox"/>					
Rights-based approach	<input type="checkbox"/>					
Disability inclusion	<input type="checkbox"/>					
Environmental sustainability	<input type="checkbox"/>					
Other (specify)....	<input type="checkbox"/>					

5. If one or several cross-cutting issues were not usually included, please explain the reasons.

6. Overall, what are the most common challenges experienced during the evaluation process? (Please mark all that apply)

- Lack of information about the intervention (M&E data, secondary data, etc.)
- Limited access to internal stakeholders
- Limited access to external stakeholders, including final beneficiaries
- The scope of the evaluation is not aligned with the available resources
- Difficulties in data collection
- Language and cultural barriers
- Reliability of available data
- Communication with IOM staff
- Lack of communication or unclear requirements from IOM staff

- Evaluation capacities and skills of the evaluator
- Communication with internal evaluators
- Communication with external evaluators
- Inadequate planning or unrealistic workplan
- Others (Please specify) \_\_\_\_\_

7. Based on your experience, to what extent do you consider that the financial resources allocated to evaluations are correlated with the quality and relevance of the final outputs?

- Completely correlated
- Somewhat correlated
- Poorly correlated
- Not correlated at all

Comments (*optional*)

8. Please rate your overall level of agreement with the following statements for the evaluations managed by you since 2020

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	N/A
Qualified and experienced <b>Internal evaluators</b> were easily identified for conducting the evaluation	<input type="checkbox"/>						
Qualified and experienced <b>External evaluators</b> were easily identified for conducting the evaluation	<input type="checkbox"/>						
<b>Time frame</b> for completing the evaluation was adequate to produce a quality product	<input type="checkbox"/>						
<b>Adequate Budget</b>	<input type="checkbox"/>						
<b>Criteria and questions</b> used for the evaluations were adequate	<input type="checkbox"/>						
Evaluation <b>methodology</b> was adequate for expressed purpose and objective	<input type="checkbox"/>						
<b>Internal communication</b> during the evaluation was fluent	<input type="checkbox"/>						

9. On average, how would you rate the quality of INTERNAL evaluations you have managed since 2020 for each of the following criteria? (Rating from 1 to 5 for each of them, where 1 signifies high quality, and 5 indicates low quality)

	1	2	3	4	5	N/A
Methodological rigor	<input type="checkbox"/>					
Relevance	<input type="checkbox"/>					
Objectivity	<input type="checkbox"/>					
Sufficient information from appropriate sources	<input type="checkbox"/>					
Utility	<input type="checkbox"/>					
Timeliness	<input type="checkbox"/>					

10. On average, how would you rate the quality of EXTERNAL evaluations you have managed since 2020 for each of the following criteria? (Rating from 1 to 5 each of them, where 1 signifies high quality, and 5 indicates low quality)

	1	2	3	4	5	N/A
Methodological rigor	<input type="checkbox"/>					
Relevance	<input type="checkbox"/>					
Objectivity	<input type="checkbox"/>					
Sufficient information from appropriate sources	<input type="checkbox"/>					
Utility	<input type="checkbox"/>					
Timeliness	<input type="checkbox"/>					

11. Based on your experience, which of the following factors has the greatest influence on the quality of the evaluation report? (Only one can be selected)

- Quality and clarity of the evaluators' recommendations.
- Engagement and participation of stakeholders during the evaluation process.
- Alignment of evaluation findings with organizational priorities and decision-making needs.
- Timing of the evaluation concerning the institutional programming cycle
- Use of the UNEG checklist
- Use of IOM Quality Control Tools terms of reference, inception reports and evaluation reports
- Others (Please specify)

## SECTION 2 – Evaluation Capacities and Skills

1. Have you received any specific evaluation training? (multiple answer)

- No
- Yes, provided by IOM
- Yes, outside the IOM

If within IOM, please select which training(s)

- Project development
- Project management
- M&E e-learning Course
- Internal Evaluator training
- Other: \_\_\_\_\_

2. If you received evaluation training provided by the IOM, how useful was it towards improving your work on evaluation?

- Very useful
- Useful
- Not very useful
- Not useful at all
- N/A

3. When involved in evaluations, how frequently do you use/consult the following IOM evaluation documents?

	Never	Rarely	Sometimes	Often	Always	Do not know the resource
<i>IOM Project Handbook</i>	<input type="checkbox"/>					
<i>IOM Evaluation Guidelines</i>	<input type="checkbox"/>					
<i>Evaluation Brief Guidance</i>	<input type="checkbox"/>					
<i>Guidance for Addressing Gender in Evaluations</i>	<input type="checkbox"/>					

	Never	Rarely	Sometimes	Often	Always	Do not know the resource
<i>Guidance for Evaluator Managers</i>	<input type="checkbox"/>					
<i>Guidance for Evaluators</i>	<input type="checkbox"/>					
<i>Guidance on quality management of evaluations</i>	<input type="checkbox"/>					
<i>Evaluation guidelines for IOM Development Fund Projects</i>	<input type="checkbox"/>					
Internal Evaluator training materials	<input type="checkbox"/>					
M&E e-learning Course materials	<input type="checkbox"/>					
Others (Please specify _____)	<input type="checkbox"/>					

4. To what extent do you consider yourself ready to conduct evaluations as part of your tasks?

	Not at all	To a small extent	To a moderate extent	To a great extent
Skilled in managing evaluations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Motivated to manage evaluations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Committed to promoting evaluation conduct and use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supported by supervisors in managing evaluations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engaged in influence others to achieve evaluative excellence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### SECTION 3 – EVALUATION IMPROVEMENT

1. What improvements could be incorporated into IOM's evaluation design, data collection, reporting and follow-up procedures?

2. From your point of view, are there any technical or methodological aspects that could contribute to improving the quality of the evaluations, as well as their findings and recommendations?

3. What suggestions do you have to improving the Quality Control Tools?

### ANNEX XII: Survey protocol evaluators (internal and external)<sup>48</sup>

#### Demographics & Control Question

1. Gender (IE & EE)

- Female
- Male
- Other gender identities

2. Evaluation related studies (IE&EE)

- No formal training on evaluation
- Evaluation training (less than 50 hours)

<sup>48</sup> Deployed as two different surveys

- Evaluation training (more than 50 hours)
- Master in Evaluation
- Evaluation related PhD

3. Years conducting evaluations (IE&EE)

- Less than 1 year
- 1 to 3 years
- 4 to 6 years
- 7 to 9 years
- 10 and more years

4. How many evaluations have you conducted for IOM? (IE&EE)

- One evaluation
- 2 - 3 evaluations
- More than 3 evaluations

5. What was your role during the evaluation conducted for IOM? (multiple choice) (IE&EE)

- Sole evaluator
- Team leader
- Team member
- Other (please specify) \_\_\_\_\_

6. Years working for IOM (IE)

- Less than 1 year
- 1 to 3 years
- 4 to 6 years
- 7 to 9 years
- 10 and more years

7. Category of your current position in IOM (IE) (*multiple selection*)

- "General"/G Staff
- "National Officer"/NO Staff
- "Professional"/P staff
- "Director"/D staff
- Service contract
- Consultant
- Intern
- Other

8. What is your experience in evaluation? (Multiple choice) (IE)

- As a PM
- As a project developer
- As a project assistant
- As an M&E assistant
- As an M&E officer
- As an M&E focal point
- Other: \_\_\_\_\_
- No experience

## SECTION 1 – EVALUATION PROCESS AND INVOLVEMENT

1. How did you get involved in the evaluation process? (IE&EE)

- Open call through the internal evaluator roster (IE)
- Open call through external evaluation platforms/networks (EE)
- Direct assignment /allocation (IE)
- Direct invitation to the bid process (EE)
- Others (please specify) \_\_\_\_\_

2. Concerning the ToR design(s), what changes were usually made during the evaluation inception phase(s)?  
(Multiple option available) – (IE&EE)

- No changes allowed
- No changes conducted
- The evaluation purpose /goal was modified.
- The evaluation scope was modified.
- Evaluation criteria were added or modified
- Evaluation questions were added or modified
- Data collection and analysis strategies modified
- Others (please specify) \_\_\_\_\_

3. To what extent do you consider that IOM guidelines, reference materials, and methodological guidelines are sufficient and clear enough to guide and support the evaluation(s) methodologically? (IE&EE)

- Very useful and clear
- Somewhat useful and clear
- Moderately useful and clear
- Not very useful and clear
- Not useful at all and clear

If used, please specify which one(s):

4. To what extent did you use the UNEG Quality Checklists and the IOM Evaluation Quality Controls Tools to improve the quality of the evaluation deliverables? (IE&EE)

- To a great extent
- Frequently used
- Rarely used
- Not used at all

5. As far as you know, to whom were the deliverables (inception report, draft evaluation reports, etc.) usually sent for comments? [Please mark all that apply]. (IE &EE)

- Staff involved in the project/intervention being evaluated
- Thematic specialists not directly involved in the implementation of the project/initiative
- Senior Management (Chief of Mission, Regional Directors, etc.)
- IOM Central Evaluation Division (OIO)
- Regional monitoring and evaluation officers
- National counterparts
- Staff from other UN entities
- INGOs, NGOs and CSOs
- Donor(s)
- No one
- Other, please specify \_\_\_\_\_
- Don't know

6. Please rate the extent to which you consider comments received improved the overall quality of the draft evaluation report(s). (IE&EE)

- Strongly agree
- Agree
- Somewhat agree
- Somewhat disagree
- Disagree
- Strongly disagree
- Usually, no comments are received

7. To what extent do you consider the process of reviewing and discussing the report(s) to be sufficiently inclusive and participatory to gather the most relevant information for the analysis? (IE&EE)

- Strongly agree
- Agree
- Somewhat agree
- Somewhat disagree
- Disagree
- Strongly disagree

8. Do you consider that the edits and suggestions received from IOM Evaluation Managers or Focal Points could have compromised the independence and impartiality of the evaluation report(s)? (IE & EE)

- Strongly agree
- Agree
- Somewhat agree
- Somewhat disagree
- Disagree
- Strongly disagree

9. What improvements could be incorporated into IOM's evaluation design, development and revision/follow-up procedures? (IE&EE)

10. Which of the following cross-cutting issues were consistently considered in the evaluation process? Tick all that apply. (IE&EE)

	Always	Usually	Sometimes	Rarely	Never	N/A
Gender mainstreaming	<input type="checkbox"/>					
Rights-based approach	<input type="checkbox"/>					
Disability inclusion	<input type="checkbox"/>					
Environmental sustainability	<input type="checkbox"/>					
Other (specify)....	<input type="checkbox"/>					

11. If one or several cross-cutting issues were not usually included, please explain the reasons. (IE&EE)

12. What were the common challenges experienced during the evaluation process, from design, implementation, and reporting to follow-up? (Multiple choice) (IE&EE)

- Lack of information about the intervention (M&E data, secondary data, etc.)
- Limited access to IOM staff
- Limited access to or response from external stakeholders, including final beneficiaries
- The scope of the evaluation is not aligned with the available resources
- Difficulties in data collection
- Language and cultural barriers
- Reliability of available data
- Communication with IOM staff
- Lack of communication or unclear requirements from IOM staff
- Inadequate planning or unrealistic workplan
- Others (Please specify) \_\_\_\_\_

13. Based on your experience, to what extent do you consider that the resources allocated to evaluations are correlated with the quality and relevance of the final outputs? (IE&EE)

- Highly correlated
- Somewhat correlated
- Poorly correlated
- Not correlated at all

14. Please rate your level of agreement with the following statements (IE&EE)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree
<b>Time frame</b> for completing evaluations was adequate	<input type="checkbox"/>					
Evaluations were <b>timely</b> conducted to be used for decision-making purposes	<input type="checkbox"/>					
<b>Criteria and questions</b> used for evaluations were adequate	<input type="checkbox"/>					
Adequate Budget	<input type="checkbox"/>					
Evaluations <b>methodology</b> was adequate for expressed purposes	<input type="checkbox"/>					
<b>Internal communication</b> during the evaluations was fluent	<input type="checkbox"/>					

## SECTION 2 – EVALUATION CAPACITIES AND SKILLS

15. Have you received any specific evaluation training? (IE)

- No
- Yes, provided by IOM
- Yes, outside the IOM

If you received specific evaluation training within IOM, please select which training(s) (IE)

- Project development
- Project management
- M&E e-learning Course
- Internal Evaluator training
- Other: \_\_\_\_\_

16. If you received evaluation training provided by the IOM, how useful was it towards improving your work on evaluation? (IE)

- Very useful
- Useful
- Not very useful
- Not useful at all
- N/A

17. When conducting evaluations, how frequently do you access and use/consult the following evaluation documents? (IE)

	Never	Rarely	Sometimes	Often	Always	Do not know the resource
<i>IOM Project Handbook</i>	<input type="checkbox"/>					
<i>IOM Evaluation Guidelines</i>	<input type="checkbox"/>					
<i>Evaluation Brief Guidance</i>	<input type="checkbox"/>					
<i>Guidance for Addressing Gender in Evaluations</i>	<input type="checkbox"/>					

	Never	Rarely	Sometimes	Often	Always	Do not know the resource
<i>Guidance for Evaluator Managers</i>	<input type="checkbox"/>					
<i>Guidance for Evaluators</i>	<input type="checkbox"/>					
<i>Guidance on quality management of evaluations</i>	<input type="checkbox"/>					
<i>Evaluation guidelines for IOM Development Fund Projects</i>	<input type="checkbox"/>					
<i>Internal Evaluator training materials</i>	<input type="checkbox"/>					
M&E e-learning Course materials	<input type="checkbox"/>					
Others (Please specify _____)	<input type="checkbox"/>					

18. How would you rate the following in relation to conducting evaluations? (IE)

	None	Little	Some	High
Your evaluation skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Your motivated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Your commitment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The support you receive from the evaluation manager(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Open to sharing the evaluation results with other colleagues and/or key partners	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### SECTION 3 – EVALUATION IMPROVEMENT

19. Based on your experience, do you have any recommendations to improve the quality of evaluations and the content of their reports? (IE & EE)?

20. Is there anything that should be done differently to enhance the dissemination, follow-up and use of evaluations? (IE & EE)?

### ANNEX XIII: Terms of reference

(excluding annexes)

#### Background

Evaluation contributes to organizational learning, accountability, adaptation and evidence-based decision-making, and to be able to operate in such manner, it needs to be credible, methodologically sound and to meet UNEG and IOM quality standards. The evaluation function in IOM operates at two levels: central evaluations covering corporate and strategic evaluations, commissioned and managed by the Central Evaluation Division (CED), and decentralized evaluations covering mainly programme and project evaluations, and usually commissioned by team leaders at country, regional and/or Headquarters levels.

The CED provides strategic direction, sets norms and standards for both central and decentralized evaluations, develops institutional policies and guidelines and offers technical support to departments and offices for the reinforcement of IOM evaluation culture. All evaluations are made publicly available

on the [IOM Evaluation repository](#) and on the [UNEG website](#), managed by CED, to promote use among UN partners and other external users.

The management of decentralized evaluations at IOM is supported by a network of Regional Officers in charge of Evaluation<sup>49</sup>, who play an important role in ensuring that decentralized evaluations are of relevant quality standards in their respective regions.

The [Multilateral Organization Performance Network \(MOPAN\) 2017-18 assessment of IOM](#), when looking at evidence-based planning and programming and at evaluation, highlighted that quality assurance systems were lacking. Consequently, IOM commissioned and conducted a [Meta-Evaluation of IOM's internal and external evaluations 2017-2019](#) assessing the overall quality of evaluations in IOM. The evaluators established criteria of quality for reviewing the evaluations and used it for the meta-evaluation in 2020. These criteria were replicated in subsequent related guidance (see below) and remain a reference for any future exercise.

In 2021, a [UNEG/OECD-DAC Peer Review of the IOM evaluation function](#) provided a series of recommendations as to how evaluation quality in IOM can be improved, some aligned to the ones of the meta-evaluation and MOPAN reports. The recommendations include: (a) introducing a quality assurance system for central and decentralized evaluations, ensuring that there is a quality assurance for terms of references, inception, draft and final evaluation reports; (b) introducing an external post-hoc quality assessment of all evaluations; and (c) encouraging the use of evaluation reference groups.

To follow up on these recommendations, a [Feasibility Study on Quality Management Mechanism](#) was conducted in 2021, including a review of systems in place in other UN agencies and a new sample of evaluation reports published in 2020 and 2021. The study emphasized the need for a checklist to review the quality of the terms of reference by the evaluation manager, and tools for the quality control during the inception and reporting phases of the evaluation. It also recommended a secondary check by the Regional M&E Officers and an annual peer or external review of selected evaluations.

In line with these recommendations, the [Guidance on Quality Management of IOM Evaluations](#) was launched in 2022, including the notions of quality assurance and quality control<sup>50</sup>. The guidance was designed for use by IOM evaluation commissioners, evaluation managers and recruited evaluators, as well as for donors, governments and other partners interested in the quality management of IOM evaluations. It aims to ensure a common understanding of the requirements related to the quality of IOM evaluations and includes a description of the mechanisms designed to review the terms of reference of evaluations and to rate the quality of inception and evaluation reports, all of them being key technical documents that can influence the overall quality of an evaluation.

During the conduct of the first Meta-evaluation in 2020, it became clear that the regular conduct of such evaluation exercises in IOM deserves to be formally established in order to allow an analysis of the quality of evaluation. The Guidance on Quality Management of IOM evaluations says: "The IOM Central Evaluation Division will periodically conduct meta-evaluations and/or peer reviews to judge the quality, merit, worth and significance of evaluations using the criteria and rating tools of the guidance, which are also in line with the rating proposed by the meta-evaluation of 2020. The meta evaluations/peer reviews may also provide further recommendations on the actions to be taken for the continued improvement of the quality of evaluations in IOM". The present exercise is part of this process.

The principles and requirements included in the Guidance highlighting the significance of quality have been consistently emphasized in the production, dissemination and use of other guidelines, for instance, the [IOM Monitoring and Evaluation Guidelines](#), the [IOM Guidance for Evaluation Managers](#), the [IOM Guidance for Evaluators](#), the [IOM Guidance on the use of evaluations and follow-up of recommendations](#), and through technical assistance and coaching, including courses, webinars and training sessions.

---

<sup>49</sup> Since 2020, this function has evolved and may still see some adjustment in the framework of the 2024 organizational restructuring. In 2020, Regional M&E Officers – P2 (ROMEO) were performing this task, replaced in 2022 by Regional M&E and Risk Management Officers -P4 (ROMERO), a position which is again under review within the restructuring of Regional Offices.

<sup>50</sup> Quality control is defined as "a part of quality management focused on fulfilling quality requirements". It is one activity related to quality assurance, which is "part of quality management focused on providing confidence that quality requirements will be fulfilled" ISO 9000

## **Objective of the evaluation**

The objective of the meta-evaluation is to assess the quality and use of internal and external evaluations (centralized and decentralized) conducted between 2020 and 2024 (five years), with the aim of providing actionable recommendations to enhance the quality and utilization of evaluations, and to assess the progress in comparison with the 2017-2019 Meta-evaluation.

The selected evaluation reports and their terms of reference (ToRs) are to be evaluated against the criteria defined in the Guidance on Quality Management of IOM evaluations, and the Meta-Evaluation of IOM's internal and external evaluations 2017-2019, both grounded in the [OECD-DAC](#) and [ALNAP](#) criteria and in the [UNEG evaluation norms, standards](#) and quality criteria<sup>51</sup>. The evaluation will provide IOM staff and partners with feedback on the quality of its evaluations, and if the institutional investment and capacity-building efforts to support the conduct and quality control of evaluations have made a change in its quality.

The main users of this evaluation will be the IOM Evaluation staff, including the Regional Officers in charge of Evaluation, who are tasked to implement the related quality management mechanisms and to provide support and guidance. This evaluation will also be relevant for IOM's senior management as a means to improve the quality of the evaluation reports to enable the Organization to use the evaluation recommendations and improve programming and decision-making. The evaluation is also relevant for Member States, donors and affected populations by providing an update on the quality of evaluations, thereby enhancing transparency and ensuring accountability.

## **Scope of the evaluation**

The meta-evaluation will collate and assess information contained in the internal and external evaluation reports (centralized and decentralized) for the period 2020 to 2024 and the ToRs and/or inception reports.<sup>4</sup> The total number covering this period from 2020 to the start date of the contract

Year	Number of Internal Evaluations	Number of External Evaluations*	Total
2020	23	49	72
2021	17	49	66
2022	17	46	63
2023	7	323	30
2024	0	3	3
<b>Total**</b>	<b>64</b>	<b>170</b>	<b>234</b>

\*Including mixed internal and external evaluations, and central evaluations as of Aug 5, 2024<sup>52</sup>

\*\* Excludes assessments and reviews.

## **Evaluation Questions**

The main evaluation questions to be addressed are:

- To what extent are IOM evaluation reports and their ToRs/Inception reports meeting the IOM and UNEG evaluation quality standards using the IOM quality assessment criteria, including for cross-cutting issues?
- On which quality aspects or factors do the evaluation processes (from preparation to reporting) and evaluation reports excel, and where are they falling short?
- To what extent is there evidence on the use of the quality assurance tools outlined in the Guidance on Quality Management of IOM evaluations?
- What can be determined about the overall quality of the evaluation reports and are there opportunities for improvement ?
- What changes can be observed in terms of quality as compared with the previous meta-evaluation, and what factors may have contributed towards these changes?

<sup>51</sup> See the UNEG [Quality Checklist for Terms of Reference and Inception Reports](#) and the [Quality Checklist for Evaluation Reports](#).

<sup>52</sup> See IOM evaluation dashboard at See dashboard at: <https://evaluation.iom.int/repository-dashboard>

- Have the cost of/resources assigned to the evaluations influenced the quality of the final product?
- To what extent are the definition of required competencies and the recruitment process of evaluators influencing quality, if the case?
- What factors facilitated or hindered the use of the tools to ensure the quality of evaluations?
- Other than the use of tools to ensure the quality of evaluations, what other factors have hindered or contributed to ensure the quality of evaluations?
- The meta-evaluation should provide some actionable recommendations for improving the use and quality of IOM's evaluations, for adjusting the existing guidelines and tools and for identifying lessons learned or good practices.

## **Methodology**

A documentation review<sup>5</sup> will be the main methodological tool used for the evaluation and will include: the first meta-evaluation, evaluation reports and their ToRs/inception reports (including annexes) published between 2020 and 2024, information related to IOM capacity building and technical assistance efforts, as well as IOM evaluation guidelines and UNEG guidance documents. The full list of evaluations conducted during the period is included in Annex 1. Most of the documents required for this evaluation are available in the evaluation repository<sup>53</sup> or under the section Technical References of IOM Evaluation Website. Nearly one-fifth of evaluations are only available in either French or Spanish, the other IOM official languages.

The document review will be complemented by interviews with a select number of IOM staff such as the Regional Officers in charge of Evaluation and technical/thematic experts to discuss the evaluation process and the support provided in the conduct of and use of evaluations and in quality setting. Furthermore, surveys will be conducted with the evaluation managers and/or programme managers of the evaluations under review, focusing on the use of the tools to ensure the quality of evaluations. The analysis of quality should also include a review of the adherence to the guidance for the production of management responses as they provide an important feedback on the quality and relevance of the evaluations.

The evaluation approach, the data collection and the methodology for analysis will be further detailed in the inception report to be produced by the recruited consultant(s). As mentioned, the analytical framework for assessing the quality of the evaluations should be in line with the Guidance on Quality Management of IOM evaluations and the Meta-Evaluation of IOM's internal and external evaluations 2017-2019. The technical management of the sampling and its size must be detailed in the technical application and the list of evaluations selected will be finalized during the inception phase. A preliminary set of evaluations is suggested in Annex 1<sup>54</sup>. The consultant(s) are required to indicate if they have participated in evaluations covered in the study to remove them from the sample period. The sample should encompass both central and decentralized evaluations, covering a variety of strategies, thematic areas, and interventions at multi-country and country levels. It should span different regions and include evaluations conducted by a diverse set of evaluators, both internal and external, across different years within the specified period.

## **Expected deliverables**

The selected company is expected to deliver the following:

- An inception report that includes an evaluation matrix and an analytical framework for evaluating the evaluations, including criteria for the selection of evaluations to be assessed;
- The analytical framework and its results;
- A draft report for comments;
- A final report;

---

<sup>53</sup> <https://evaluation.iom.int/evaluation-search-pdf>

<sup>54</sup> A preliminary sample of 110 evaluations is suggested out of the 234. The evaluator can modify the sample, while ensuring the inclusion of a diverse set of evaluations in terms of evaluators/authors, type of evaluator, thematic areas, countries covered and level (centralized vs decentralized).

- A summary report or brief (2-4 pages), presenting the evaluation, key findings, lessons learned, good practices and the key actionable recommendations; and
- A final presentation of the evaluation results.

### **Roles and responsibilities**

**Management:** The CED will manage the meta-evaluation and will act as the primary focal point for the selected evaluator(s). CED will provide the evaluator(s) with the list of documents, and suggested interviewees. The Management will also organize a kick-off meeting with the evaluator(s) once selected and will be responsible for the final approval of all the deliverables. The CED will be responsible for completing the Management Response.

**Reference group (RG):** A reference group will be created to provide feedback to the evaluator(s) during the meta-evaluation, in particular by reviewing the samples (reports and interviews/surveys) and deliverables. The reference group will review the lessons learned, good practices and actionable recommendations identified by the evaluator(s) having in mind the importance of ensuring that they can be adequately used and implemented. Members of the RG should be involved in the promotion and/or use of IOM Quality Management guidance, such the Regional Officers in charge of Evaluation.

**Evaluator(s):** The selected company will be responsible for completing the evaluation in accordance with the present ToRs, inception report and the deliverables set out therein. All deliverables will need to be approved by the RG and Management.

### **Evaluation Timeline**

Following the selection of the company, the following *indicative* timeline is expected:

<b>Activity</b>	<b>Days</b>	<b>Timeframe</b>
Inception report, analytical framework and data collection tools	10	August 2024
Data collection and analysis	23	September – October 2024
Drafting report	8	November 2024
Finalization of report, summary report or brief and presentation	5	November - December 2024
<b>TOTAL DAYS FOR CONSULTANCY</b>	<b>46</b>	

The work is expected to be done remotely, so the location of the company is not relevant. No travel is required.

### **Payment schedule**

The services will be paid as follows:

- 30% upon reception, verification and successful completion of the inception report
- 30% upon reception, verification and successful completion of the draft final report
- 40% upon reception, verification and successful completion of the final report

## **ANNEX XIV: Inception report**

(Excluding Annexes)

### **BACKGROUND**

#### **Introduction**

This report documents the results of the inception phase and activities directed to scope the Meta-Evaluation of the International Organization for Migration (IOM) for the evaluations conducted from 2020 to 2024 (from now on, “the Meta-Evaluation”).

The report outlines the context that makes the Meta-Evaluation pertinent. It details the aims, scope, and methodology, including the evaluation and research questions, evaluation techniques, sampling strategy, and approaches for data gathering, analysis, and interpretation.

The inception report and the Terms of Reference (ToR) will direct the consultancy, which will be carried out by an external consulting team from Artival Research & Evaluation. The consultancy is scheduled from November 2024 to March 2025. The IOM and other pertinent partners will review the submitted materials, offer feedback, and approve the final report.

#### **About the object of the Meta-Evaluation**

The ToR states that evaluation is important for learning, accountability, adaptation, and evidence-based decision-making. The IOM's evaluation function works on two levels: central evaluations, which encompass corporate and strategic evaluations and are overseen and managed by the Central Evaluation Division (CED), and decentralized evaluations, which primarily focus on programme and project evaluations and are typically commissioned by team leaders at the country, regional, and/or Headquarters levels.

After the Multilateral Organization Performance Network (MOPAN) 2017-18 assessment of IOM, the Organization conducted the first Meta-Evaluation of its internal and external evaluations from 2017 to 2019. The resulting Meta-Evaluation, presented in 2020, stated that the regular conduct of such Meta-Evaluation exercises in IOM deserves to be formally established to allow an analysis of the quality of evaluation.

This exercise and the UNEG/OECD-DAC Peer Review of the IOM evaluation function (2021) concluded with recommendations to improve the quality of IOM evaluations. The recommendations include: (a) introducing a quality assurance system for central and decentralized evaluations, ensuring that there is quality assurance for terms of references, inception, draft, and final evaluation reports; (b) introducing an external post-hoc quality assessment of all evaluations; and (c) encouraging the use of evaluation reference groups.

The IOM has been fostering a culture of evaluative understanding and practice to enhance the quality assurance system. This is promoted by considering the growing recognition of evaluations' importance in reporting intervention outcomes and lessons learned and pinpointing innovative and replicable practices.

Since 2020, the IOM efforts for strengthening the organizational evaluation function included a feasibility study on a Quality Management Mechanism (2021), a review of systems in place in other UN agencies, and a new sample of evaluation reports published in 2020 and 2021. To ensure that all quality requirements are met, several guidelines were approved—Guidance for Evaluator Managers (2021), Guidance for Evaluators (2021), Guidance on Quality Management of IOM Evaluations (2022), and Guidance on the Use of Evaluations and Follow-up of Evaluation Recommendations (2022)—as well as IOM Evaluation Strategies (2021-2023 and 2024-2028). A review of the use and follow-up of management responses in IOM is ongoing and expected to be completed at the end of 2024.

In this context, the Meta-Evaluation will assess the quality of the annual evaluation reports (analysis of the evolution), the type of evaluator, and the internal use of the quality assurance tools for the evaluations conducted by IOM between 2020 and 2024.

## OBJECTIVES AND SCOPE

### Objectives

In line with the ToR, the **main objectives** of the Meta-Evaluation are:

To assess the quality of internal and external evaluations (centralized and decentralized) conducted between 2020 and 2024 and provide actionable recommendations to enhance the quality and utilization of evaluations.

Assess the relevance and use of the quality assessment tool introduced, analysing opportunities and limitations on its use and identifying other relevant factors that influence the use of evaluation results.

To assess the evolution of the evaluation quality compared with the 2017-2019 Meta-Evaluation.

### Scope

In line with the ToR, the **scope of the Meta-Evaluation** will assess internal and external evaluation reports (centralized and decentralized) for 2020 – 2024 and, to a certain extent, the ToRs associated with these evaluations. Additionally, the tool developed for assessing the inception reports will be tested. A total of 90 evaluation reports and 20 ToR documents will be assessed. The geographical coverage includes all the regions where IOM operates. More information about the sampling strategy and the evaluation reports to be assessed can be found in section 3.4 (Figure 1).

The **Meta-Evaluation will focus on maximising the utility of its results and their actual use**. To this end, the design and process will carefully consider meeting the information needs of primary and secondary audiences, as tentatively expressed in Table 1.

Table 1. Intended users and expected use for the Meta-Evaluation

Type of User	Designation	Intended use
PRIMARY	Central Evaluation Division Regional/Country level Planning and M&E Officers	Strengthening of IOM's evaluation function. Improve preparation, implementation, and resource management for evaluation. Accountability Learning
	IOM's senior management and IOM in general	To improve the quality and use of its evaluation reports. To improve practices in the future as IOM staff manage, conduct, or use evaluations
SECONDARY	Other UN agencies and organizations	Organizational learning/development
	Beneficiaries of IOM interventions / civil society	Learning/development, accountability
	Member States and donors	Accountability

The **primary users** of this evaluation will be the IOM Evaluation staff, including the Regional and Country Officers responsible for evaluation, who are tasked with implementing quality management mechanisms and offering support and guidance. Additionally, this evaluation will benefit IOM's senior management by enhancing the quality of evaluation reports, allowing the organization to use the evaluation recommendations to improve programming and decision-making. It is also significant for other UN agencies and organizations, civil society, Member States, donors, and the beneficiaries of IOM interventions by providing insights into the quality of evaluations, thereby improving transparency and ensuring accountability.

Engagement with the primary intended users will extend beyond merely providing information; it will encompass strategies and approaches for effectively implementing and communicating findings, conclusions, and recommendations. This analysis will help the consulting team craft specific, realistic, and pertinent recommendations grounded in previous experiences, a participatory process with key stakeholders, and a clear, logical, evidence-based link among findings, conclusions, and recommendations.

## APPROACH AND METHODOLOGY

This section establishes the approach and methodology to guide the data collection, analysis, interpretation, and framing of the final deliverables.

### Approach

The Meta-Evaluation will adopt a utilisation-focused approach and be executed using a mixed-method concurrent triangulation research strategy. The consulting team will simultaneously gather quantitative and qualitative data and then compare these datasets to identify convergences, divergences, or a combination of both. This process will involve confirmation, disconfirmation, cross-validation, and corroboration to derive key findings that will inform conclusions and recommendations. In practice, quantitative data will highlight statistical trends, while qualitative data collected during the assessment will offer explanations and nuances to these trends.

The **driving questions** for the Meta-Evaluation can be summarized as “What is the quality of the evaluations conducted, how has evaluation quality evolved since 2020, and what is required to strengthen IOM’s evaluation function in the future?” The following table reveals the key concerns and the different areas considered for the analysis.

Table 2. Key areas of analysis of the Meta-Evaluation

Key concern	Areas of analysis
Evaluation quality	Overall evaluation quality and ways to improve it Strengths and weaknesses of evaluation processes Influence of cost and timing on the overall quality of evaluation Evolution in the quality of evaluation since 2017
Evaluation capacities	Evaluation quality per type of evaluator (internal and external)
Use of evaluation tools	Use of the quality assurance tools and testing the inception report quality assessment tool Issues that facilitated or hindered the use of evaluation tools Factors that influenced the use of evaluation results

### Methodology

The following **methodological strategies** and data-gathering techniques will be considered:

Table 3. Methodological approach

Methodological strategies	Research Techniques	Data gathering tools
Qualitative strategies (Discourse analysis)	Semi-structured interviews Document review	Interview protocols Scoring tool
Quantitative strategies (Descriptive and inferential statistical analysis)	Analytical framework (desk review) Online survey	Questionnaires Scoring tool

The methodological proposal intends to solve methodological challenges by:

- **Combining quantitative and qualitative data analysis:** Quantitative data gathered by primary and secondary sources (such as the scoring tool, the online survey, and document reviews) can provide substantial information on effectiveness, while data gathered through qualitative methods (interviews, document reviews, and the scoring tool's open-ended field) can provide meaningful information about all components and notably on the utility and relevance of the evaluations and ToR produced.

- **Using different and complementary sources of information:** The proposed approach uses different sources to fill the gap between the absence of information, non-updated information, and unreliable information. Data obtained from the assessment of reports, ToR, and inception reports will be complemented by individual opinions (structured interviews and surveys). Different sources will allow for triangulation of information, especially when searching for good practices and recommendations for the future.

**The four complementary research techniques considered are:**

- **Assessment with the scoring tool<sup>55</sup>:** The analytical framework and its scoring tool will provide essential data on which to base the Meta-Evaluation results, allowing the emergence of statistical trends utilising proportions, measures of central tendency, and statistical association through correlation coefficients. The statistical association analysis will determine the extent to which evaluation budgets and the type of agent conducting the evaluation (internal/external) correlate to the ratings regarding evaluation quality. Once the quantitative database is ready, specific software will be used to conduct the analysis and produce the visuals (SPSS, Python, Tableau, and/or Excel). The tool will also gather qualitative information about the evaluators' perceptions that will go beyond the rating given. This will be done through the comments made to each rating.
- **Assessment of the quality assessment tool for inception reports:** Based on the Meta-Evaluation developed by Artival in the first Meta-Evaluation exercise, CED produced a tool to assess the quality of inception reports. The tool will be assessed during this consultancy (three assessments will be considered) to provide feedback on its adequacy. The team will also assess the adequacy of the adapted Meta-Evaluation and ToR assessment tools developed by CED.
- **Semi-structured interviews:** Interviews with a select number of IOM staff will be the other crucial method in the Meta-Evaluation, as they will allow the triangulation of information, points of view, and opinions of different stakeholders. The views and information provided will deepen the findings of the analysis. Around 15 interviews are expected to take place, targeting country and regional monitoring and evaluation officers and a selection of key informants suggested by CED and the reference group.
- **Survey:** An online survey was developed and will be launched to complement qualitative information with related quantitative data. The questionnaires will be sent to internal and external evaluators and the managers of the evaluations under review.
- The consulting team will develop the tools concurrently. These are informed by the key Meta-Evaluation questions and the key areas under assessment included in the Scoring tool.

Table 4. Meta-Evaluation questions<sup>56</sup> and techniques

Meta-Evaluation questions	Concept	Research Tools
<b>To what extent do IOM evaluation reports and their ToRs meet the IOM and UNEG evaluation quality standards as per the IOM quality assessment criteria?</b>	Quality of evaluations	Scoring tool Interview
On which quality aspects or factors do the evaluation processes (from preparation to reporting) and evaluation reports excel and where are they falling short?	Quality of evaluations	Scoring tool Interview Survey
To what extent is there evidence on the use of the quality assurance tools outlined in the Guidance on Quality Management of IOM evaluations?	Internal use of evaluation tools	Scoring tool Interview Survey

<sup>55</sup> The scoring tool from the initial Meta-Evaluation has been revised to better capture the specific goals of this assessment. Details of these revisions can be found in Annexes 1, 2, and 3.

<sup>56</sup> Some of the original Meta-Evaluation questions from the ToR have been consolidated to make them more straightforward and streamlined, ensuring that all the main aspects are still included.

Meta-Evaluation questions	Concept	Research Tools
What factors facilitated or hindered the use of the tools to ensure the quality of evaluations?		
What is the adequacy of the Quality Assurance (QA) tools developed by CED?		
What other factors influence the use of evaluation results?		
What changes can be observed in terms of quality as compared with the previous Meta-Evaluation, and what factors (beyond the use of QA tools) may have contributed towards these changes?	Evolution of the quality of evaluation reports	Scoring tool Interviews
<b>Have the cost of/resources assigned to the evaluations positively correlated with the quality of the final product?</b>	Quality of evaluations	Scoring tool Interview Survey
<b>What are the main trends in evaluation quality identified between internal and external evaluators?</b>	Quality vs type of evaluators	Scoring tool Interview Survey

Table 4 shows that the scoring tool will play a major role in the analysis. Developed for the previous Meta-Evaluation, it has been revised to better address the specificities of this assignment. It's based on the evaluation questions, and a series of rubrics and ratings specifically developed to assess the extent to which each concept/evaluation question is addressed.

The analytical framework that will serve as the key tool to conduct the Meta-Evaluation considers **four different levels of analysis**:

**Dimensions:** The first level of analysis includes three categories, as described in Table 5.

**Components:** At the second level of analysis, there are the key sections on which the assessment of each dimension is based (e.g., evaluation background, methodology, findings, recommendations, etc.). Per the evaluation report, each component is assigned a specific weight for the overall assessment.

**Items:** Represent the general criteria for adequately assessing compliance with each component; they indicate key concepts to be considered under each component.

**Elements:** Represent the ultimate sub-criteria on which the assessment is based. One item can comprise one or more elements, and the assessment is based on whether each item complies with the set of elements considered.

Table 5. Dimensions under analysis through the Meta-Evaluation tool

DIMENSION	UTILITY
Quality of evaluation reports	Collect information about the quality of evaluation reports. Rate the quality and relevance of the information within reports.
Comprehensiveness of evaluation reports	Summary and checklist of the information contained in the evaluation reports.
Comprehensiveness of ToR	Assess the key sections and information contained in the terms of reference and inception reports.

Dimension 1 is the most comprehensive and provides useful information for analysing ten components: the executive summary, intervention background, evaluation background, questions and criteria, evaluation methods, findings, conclusions, recommendations, lessons learned, and emerging good practices (see annex 1 for more details). Each main component comprises one or more elements or criteria to base the assessment. Table 6 reveals the items and elements considered for each component's recommendations.

Table 6. Items and elements consider for component "Recommendations" in Dimension 1

COMPONENT 8: RECOMMENDATIONS – Weight 20%	
ITEMS	ELEMENTS
R1. Clarity	1. Are clear and concise 2. Are based on findings and conclusions of the report
R2. Relevance	1. Address key issues and are useful, tied to the object and purposes of the evaluation. 2. Address cross-cutting issues
R3. Responsibility	1. Specify who is called upon to act 2. Identify means for achievement 3. Specify priority or importance
R4. Actionability	1. Indicate the specific courses of action needed to remedy/or continue with the current situation 2. Provide a time frame

Once all elements under a particular component are considered, rubrics will be used to determine compliance with established criteria. The next section provides more details.

Finally, **the analytical framework includes three different scoring systems**, which are:

For Dimension 1 on the quality of evaluation reports (Annex 1), a six-point scale representing the main scoring system is included. The different rubrics are expressed in Table 7. This is an ordinal scale, so the analysis would consider frequencies, proportions, modes, and medians but not means. **Finally, a global rating will be provided for each report**, considering the ratings given and the relative weight considered for each of the components (see Figure 1).

For Dimensions 2 on the comprehensiveness of evaluation reports (Annex 2), and Dimension 3 on the comprehensiveness of the ToR (Annex 3) include dichotomous nominal scales. Therefore, the analysis will mainly consider frequencies and proportions.

Table 7. Six-point scale considered for Dimension 1 on quality of evaluation reports

0	1	2	3	4	5
Highly unsatisfactory None of the required elements are present.	Unsatisfactory Not all the elements are present, and at least one of the elements presents is incomplete.	Somewhat unsatisfactory Not all the elements are present, but those present are complete.	Somewhat satisfactory All the elements are present but not all of them are complete.	Satisfactory All the elements are present and complete.	Highly satisfactory All elements are present, complete, interconnected and the report excels in covering the item.

**Concerning the specific tools proposed for the analysis**, the consulting team proposes elaborating a map of codes, enabling a greater understanding of the information gathered. Qualitative data will be structured by component and Meta-Evaluation criteria. Each interview will be reviewed, and key information will be coded into categories and subcategories. Key categories will be identical for every data-gathering tool, allowing the comparisons and identification of trends across thematic areas and research techniques. All qualitative information will be compiled, consolidated, collated, coded, and analysed using specific software for conducting qualitative and mixed-methods research (e.g., Dedoose).

Figure 1- Relative weight for the rating of each component

			Criteria & questions 7,5%	Evaluation methods 7,5%
			Executive summary 5%	Evaluation background 5%
Findings 25%	Conclusions 20%	Recommendations 20%	Project background 5%	Lessons learned 2,5%      Good practices 2,5%

**In terms of enhancing the validity** of the evaluation, the combination of proposed methods will allow the triangulation of information, strengthening the study's validity. The consulting team will consider both data triangulation (use of a variety of data sources), method triangulation (use of a variety of methods), and investigator triangulation (use of different researchers or evaluators). Measurement validity (the extent to which evaluation indicators capture the concept of interest) will be revised and presented to key evaluation audiences for validation.

The consulting team will conduct several **reliability procedures** to ensure the reliability of the data-gathering tools. The scoring tool will be piloted in one evaluation report to establish a common rating system across all team members and enhance inter-rater reliability. Specifically, the scores given by each team member and the reasoned exposition of each component will be compared to identify discrepancies in the interpretation of criteria and rubrics. This process will facilitate the development of a standardised assessment framework, enabling consistent and comparable ratings among elements, items, and components. Once all the interventions are reviewed, the extreme cases (evaluations with the highest or lowest ratings) will be double-checked through a peer review process to ensure consistency.

The contrast between the scores assigned by each team member and the detailed explanation of each item will help identify discrepancies in the interpretation of criteria and scales. This process will facilitate aligning a common system of evaluation guidelines based on the established rubrics. Such a revision will initiate a methodological discussion regarding the boundaries and shared definitions of the concepts outlined by the rubrics.

For the qualitative analysis, the team members will ensure that the definition of codes and subcategories for qualitative data is clear and that all members understand each key concept similarly. This will be accomplished by clearly defining each code, regularly comparing data with codes, and cross-checking codes developed by different team members.

During the start of the data collection process, another round of verification will take place to ensure that the understanding of the Meta-Evaluation tool's components, items, and elements remains the same across the coders. Additionally, the Meta-Evaluation report will highlight how the consulting team ensured reliability through the data collection phase and coding of the evaluation reports. This will be done by including a document (Annex IV: Strategies to ensure consistency) comprising all the discrepancies faced during the Meta-Evaluation process and the strategies and agreements reached to address them.

## Data sources

Primary data comprises the information produced through the analytical framework and assessment of the quality assessment tool developed by CED, as well as the KII, and online survey designed. Secondary data is considered all the information on which the rating and analysis are based, as described in Table 8.

Table 8. Secondary data sources

List of key documents	Type of information collected
<b>Normative and operational framework:</b> IOM's evaluation policies and guidelines, UNEG reference documents, and toolkits and other related documentation (EPI, MOPAM, etc.)	Contextualisation and deep understanding of key frameworks and strategies considered within IOM.
<b>The evaluation reports and ToRs</b>	Key information in which to base the assessment.

### Sampling procedures

In this Meta-Evaluation, the initial pool consists of 265 evaluations. After excluding those conducted by Artival, 260 evaluations remain. To manage this large number of evaluations, a sampling strategy was employed to ensure that the sample reflected the same distribution as the original population. A sample size of 90 evaluations was selected.

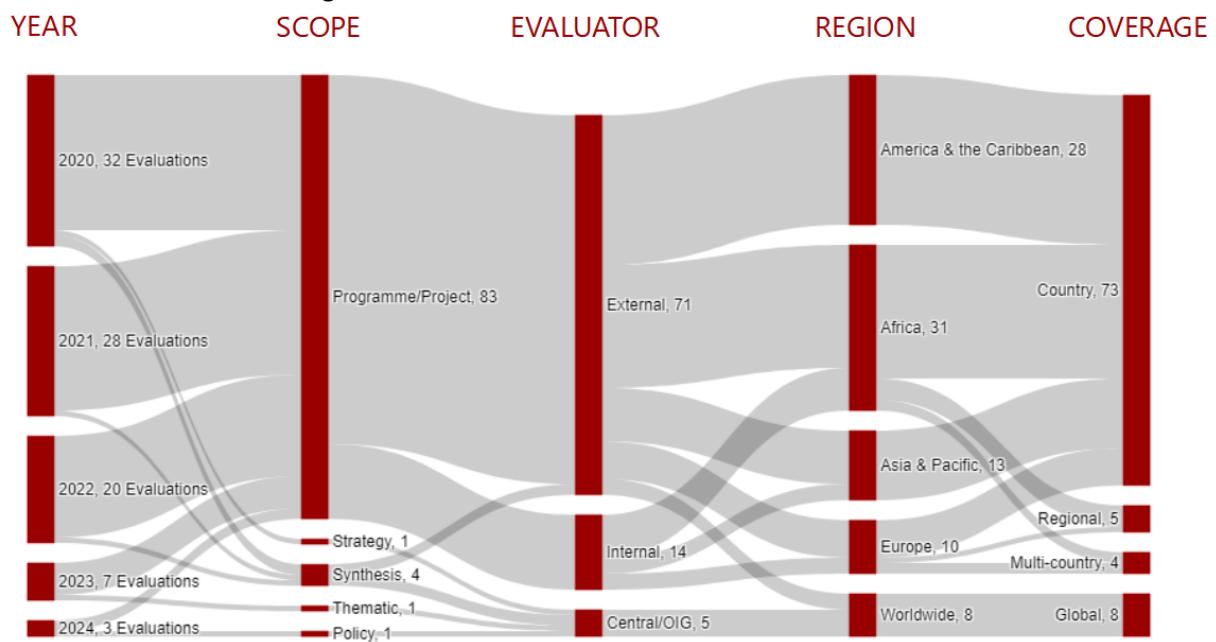
In the initial step, four key criteria were chosen to ensure the sample's validity, focusing on the main aspects crucial for evaluating the assessments: Year, Type of Evaluator, Evaluation Coverage, and Regions Covered. The objective is to consider these four criteria together rather than in sequence. This approach prevents any individual criterion from overshadowing others, which could lead to the underrepresentation of significant categories in the final sample. Combining these criteria created a measure of each value's weight relative to the total population. These weights were then used as probabilities in a Python-based random sampling algorithm, resulting in a sample of 90 evaluations.

In the second step, two additional criteria—Type of Evaluation Scope and Donor—were selected to verify the balance of the sample. Based on these criteria, the selected sample was compared with the total population to ensure that the most relevant categories related to the Type of Evaluation Scope and Donor were adequately represented in the final sample.

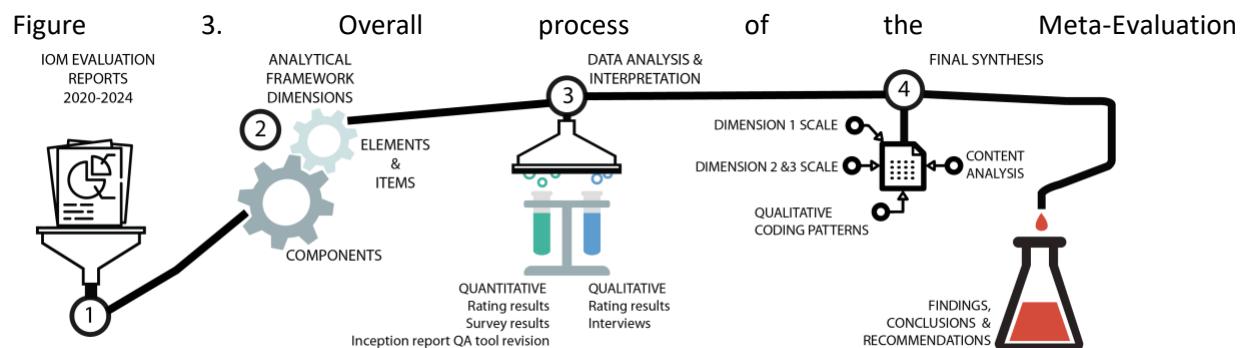
Some categories from the selected criteria were omitted during the sampling process. This is expected, as the primary goal is to reduce the number of evaluations while maintaining a representative sample. For the Regions Covered and Evaluation Scope criteria, the categories excluded from the final evaluation list are those combined, such as "RO Cairo, RO Dakar" under Regions Covered and "Strategy, Thematic" under Evaluation Scope. The final sample is considered valid since evaluations containing these categories individually are included. As a result, approximately 3% of cases for Regions Covered and about 5% for Evaluation Scope lack representation in the final sample. Additionally, for the Donor criteria, some donors are absent from the final sample due to their wide variety. Overall, only 5% of donors are not represented, with examples like the "Government of the Netherlands" and "Government of Norway," each having 3 out of the initial 260 evaluations.

From the 90 selected evaluations, a sample of 20 ToRs will be analysed (10 from the evaluations with the highest scoring and 10 with the lowest scoring) will be analysed to determine their quality based on the existing quality assurance tools. Finally, 3 Inception Reports will be analysed to pilot the use of the tool.

Figure 2. Overview of evaluations under assessment



The following figure represents the **overall process for the assessment**, from the final universe of evaluation reports under consideration to the findings, conclusions, and recommendations of the final report. The draft report will be circulated to stakeholders in the third phase of the Meta-Evaluation. This will allow each partner to be involved in the evaluation and improve the initial draft overall by completing, adjusting, or qualifying the whole evaluation and deepening the lessons learned and recommendations.



### Cross-cutting issues

Gender, human rights, disability, and environmental integration are the cross-cutting issues for the Meta-Evaluation. The assessment will determine how human rights, gender equality, disability inclusion, and environmental aspects are considered when designing evaluations and defining findings, conclusions, and recommendations.

In the evaluation methods, there is a specific aspect concerning the extent to which cross-cutting issues were considered during the evaluation (EM5), in the findings (F4), and in the relevance of recommendations (R2). The evaluation approach and methods of data collection and analysis will also be analysed to determine whether they are rights-based and gender-responsive and whether data is

disaggregated by social criteria (e.g., sex, ethnicity, age, disability, geographic location, income, or education).

### **Norms and standards**

The consulting team will ensure that the Meta-Evaluation process respects the UNEG's basic ethical and professional behaviour principles. It will follow relevant IOM policies and UNEG norms and standards<sup>57</sup>. The consultants will be fair and assess with integrity and honesty. They declare no conflicts of interest, as they have not directly or indirectly participated in any of the interventions under review or have a personal relationship with managers and/or consultants involved in those.

### **LIMITATIONS OF THE META-EVALUATION**

The assessment team has identified the following risks and limitations for the Meta-Evaluation:

Table 9. Potential risks/limitations and mitigation strategies

RISK/LIMITATION	EXPLANATION	MITIGATING STRATEGY
Limited number of evaluations with information on budget assigned and training received by internal evaluator.	Two key questions of the Meta-Evaluation aim at establishing differences on the quality of evaluation reports by budget utilised and the training received by internal evaluator.	The team will include an analysis of the results of those evaluations with information on costs and trained/non-trained internal evaluators. The relevance of the analysis will be based on the representativeness of the results.
Different types of evaluations	The IOM evaluations can be conducted internally or externally. This could lead to substantial differences within the reports that could affect their quality.	The Meta-Evaluation will consider this dual nature and will spot any difference concerning this typology.
Large range of migration thematic areas and sectors	The evaluation reports analysed in the Meta-Evaluation include interventions of many thematic areas and sectors that can be implemented and evaluated with many different methodologies.	The Meta-Evaluation will consider this and stress any challenges found for particular areas as well as mitigation strategies.
Incomplete data sets	The consulting team will require reviewing documentation including ToR, evaluation reports, inception reports and other relevant materials (resources and budgets). However, there might be limited information for some of the projects/programmes involved.	The consulting team will inform about the missing data or documentation required, and if the sample does not provide the required number of ToRs, budget, and IRs, some evaluation of the sample may be replaced.
Interviews and surveys limitations	Some of the evaluations were carried out by externals, others by internal staff who are no longer in the organization, so access to them might not be possible.	In case this limitation arises, the consulting team will try to contact those staff members or external teams who have participated in the evaluation.
Delays in receiving data/feedback	Should there be delays in receiving feedback from the relevant stakeholders on expected deliverables, the consulting team will still require the number of weeks indicated in the timeline to respond to feedback.	The consulting team proposed a specific period for stakeholders to consolidate feedback during the different stages of the consultancy. Should there be significant delays in receiving consolidated feedback, the team will work with the evaluation

<sup>57</sup> Click to access the UNEG standards.

RISK/LIMITATION	EXPLANATION	MITIGATING STRATEGY
		manager to reschedule key milestones as required. This is reflected in the updated timeframe for the inception phase.
Inconsistencies in evaluation quality results because of the QA tools developed	The introduced Meta-Evaluation and ToR quality assessment tools follow pretty much the same logic as the original scoring tool produced for the first Meta-Evaluation. However, some changes were considered in the weight of the different sections and other elements, which can have a negative impact on studying the evolution of evaluation quality over time.	If relevant inconsistencies are identified, the consulting team will ponder the effects those might have in the study of the evolution of evaluation quality through time. If necessary, adjustments to the weight system will be considered for standardisation of tools and results.

## SCHEDULE AND DELIVERABLES

### Timeframe and work plan

The Meta-Evaluation will be implemented from the 23<sup>rd</sup> of October 2024 until the 14<sup>th</sup> of April 2025.

**The consulting team will be engaged for a total of 50 working days.**

Table 10. Key deliverables and milestones

Tasks	Timeframe	Product
Contract formalisation	22nd of October 2024	Contract
Initial briefing with the Meta-Evaluation manager	7th of November 2024	Initial notes
Refinement of methods, tools and questions	18th of November 2024	Draft Inception Report
IOM comments on the Draft Inception Report	2nd of December 2024	Comments
Validation meeting	9th of December 2024	Final Inception report
Data screening and synthesis	31st of January 2025	-
Draft Final Report	17th March 2025	Draft Final Report
IOM comments on the Draft Final Report	31st March 2025	Comments
Final Report and presentation of the evaluation reports	14th April 2025	Final Report, Summary Report and Presentation

### Deliverables

- **An inception report** including a draft report outline, a detailed methodology, and a work plan. This should include, among others, a description of the key questions to be addressed, the analytical approach to be taken (with details regarding the aspects and topics that will be addressed quantitatively and/or qualitatively), and the means to address the risk of bias and quality control. The approach considers the specificities of evaluating international development and humanitarian action interventions. The CED and the Reference Group will be invited to review and comment on the inception report, and the consultants will respond in writing to all received comments. IOM must approve the inception report before work can proceed.
- **A Draft Final Report** based on the approved inception report. The draft final report will present the analysis of the selected documents and key findings and include an executive summary. It will be written in English and no longer than 50 pages, including the executive summary, excluding the annexes. The draft report will be presented to IOM and key stakeholders for comments, and the consultants will respond in writing to all comments received. IOM will conduct a quality assurance/peer review of the draft version of the report.
- **A consolidated Final Report** will present the completed analysis and respond appropriately to comments and feedback from key IOM officials. The Final Report will have a high-quality executive