

FINAL REPORT

Meta-Evaluation of IOM's internal
and external evaluations
(2017 – 2019)

April 2020



Consultancy: A Meta-Evaluation of IOM's Internal and External Evaluations

Timeframe of the interventions studied: January 2017 to December 2019

Locations (country, region, etc.) of the interventions under review: Worldwide

Consulting team: Salvador Bustamante (Coordinator), Leticia Bendelac, Jorge García and Juncal Gilsanz.

Organisation commissioning the consultancy: International Organization for Migration

Name of contact point for the consultancy: Diana Cartier

Consultancy budget: 29,650€

The Artival team acknowledges the valuable guidance and support provided by the Meta-evaluation Reference Group and the consultancy focal point, Diana Cartier. The team would also like to thank all IOM staff involved in the process, for their participation and insights.

Contact details

Artival Research & Evaluation
info@artival.es
www.artival.es

FOREWORD

IOM is pleased to share the results of the meta-evaluation undertaken in response to the comprehensive assessment report on IOM published by the Multilateral Organisation Performance Assessment Network (MOPAN).

MOPAN considered that the policy architecture for evaluation required further improvement, including through the development of a more rigorous quality assurance system. It noted that the evaluative evidence base to inform planning and design is constrained within IOM and that a stronger evaluation culture is still required. It also recommended developing a framework to enhance coverage of evaluations across the Organization and to promote the better use of evaluations in IOM decision making processes.

The meta-evaluation was commissioned by IOM's Evaluation office and conducted by an external and independent consulting firm, Artival Research and Evaluation. The evaluation assesses the quality of all individual internal and external evaluations published between 2017-2019 (88 in total), reviewing them through a very detailed set of criteria and a rigorous methodology that also includes surveys with the main stakeholders. It also reviews IOM's decision-making processes concerning programme evaluation, the capacity of IOM managers and other staff to conduct internal evaluations as well as the dissemination and use of evaluations within IOM. Finally, it provides a number of concrete recommendations for the way forward.

I am very pleased to note that the meta-evaluation is generally positive in relation to the overall quality of the evaluations conducted in IOM. Findings from this meta-evaluation are particularly encouraging given the often-limited financial resources allocated to evaluations in IOM.

The report also examines the current processes and the involvement of the different actors within IOM who should play a role in the setting-up of a stronger evaluation system, including the roles of the Evaluation function in the Office of the Inspector General, the Regional Monitoring and Evaluation (M&E) Officers and IOM staff, in particular those conducting internal evaluations. The report praises IOM's staff coaching opportunities and the capacity building initiatives developed that have been undertaken in the last years under the leadership of the Evaluation office.

The report also underlines the importance of engaging in the evaluation process with senior managers, including Directors and Chiefs of Missions, to ensure that recommendations from individual evaluations are more broadly disseminated and followed up within the Organisation and used to identify programme delivery improvements and good practices.

The recommendations of the meta-evaluation will contribute to broader efforts IOM is undertaking to improve the Organisation's efficiency and effectiveness in programme delivery; its agility and responsiveness; as well as its accountability and transparency through the application of its Internal Governance Framework (IGF). The IGF is also intended to enhance IOM's results-based management capacity, and its ability to learn and identify good practices promoting a culture of continuous improvement within the Organisation.

Further steps towards a more robust evaluation system within IOM will continue through an upcoming Peer Review of IOM's Evaluation function and policy and its adherence to the UN Evaluation Norms and Standards. This Peer Review will be managed jointly by the UN Evaluation Group and the OECD/DAC and will be published in the first quarter of 2021.

I would like to thank IOM's Evaluation office for this informative and timely initiative and reaffirm the Senior Management commitment to continue strengthening IOM's evaluation capacity in the future.

Laura THOMPSON
IOM Deputy Director General

CONTENTS

LIST OF FIGURES	5
LIST OF TABLES.....	5
ACRONYMS	7
EXECUTIVE SUMMARY.....	8
BACKGROUND.....	16
Introduction	16
About the object of the meta-evaluation.....	16
OBJECTIVES AND SCOPE	16
Objectives	16
Scope.....	17
APPROACH AND METHODOLOGY.....	17
Approach.....	18
Methodology.....	18
Data sources	22
LIMITATIONS OF THE META-EVALUATION	24
FINDINGS ON EVALUATION REPORT QUALITY.....	25
FINDINGS ON EVALUATION PROCESSES.....	45
FINDINGS ON EVALUATION CAPACITIES AND SKILLS	50
FINDINGS ONEVALUATION PROMOTION AND USE	54
CONCLUSIONS	57
GOOD PRACTICES	62
LESSONS LEARNED.....	62
RECOMMENDATIONS.....	63
LIST OF ANNEXES.....	66
ANNEX 1 – Meta-Evaluation tool - Quality of evaluation reports.....	67
ANNEX 2 – Meta-Evaluation tool - Comprehensiveness of evaluation reports.....	73
ANNEX 3 – Meta-Evaluation tool - Comprehensiveness of Terms Of Reference.....	74
ANNEX 4 – Questionnaire for Evaluation Managers	76
ANNEX 5 – Questionnaire for Internal Evaluators	83
ANNEX 6 – Questionnaire for External evaluators	89
ANNEX 7 – Interview Protocol	93
ANNEX 8 – List of documents reviewed.....	95
ANNEX 9 – List of evaluation reports.....	96
ANNEX 10 – Consistency issues considered.....	102
ANNEX 11 – List of people interviewed.....	108
ANNEX 12 - Detailed analysis per geographical area.....	109
ANNEX 13 - Inception report.....	111

LIST OF FIGURES

FIGURE 1- RELATIVE WEIGHT FOR THE RATING OF EACH COMPONENT	22
FIGURE 2. OVERVIEW OF EVALUATIONS UNDER ASSESSMENT	23
FIGURE 3. OVERALL PROCESS OF THE META-EVALUATION	24
FIGURE 4. EVOLUTION OF GLOBAL RATING / DISTRIBUTION OF REPORTS PER RATING.	26
FIGURE 5. GLOBAL RATINGS PER GEOGRAPHICAL AREA.....	27
FIGURE 6. GLOBAL MEDIAN RATINGS PER AREA/REGION	27
FIGURE 7. EVOLUTION OF MEDIAN RATING PER REGION	27
FIGURE 8. GLOBAL MEDIAN RATING PER THEMATIC SECTOR/AREA.....	29
FIGURE 9. CORRELATION MODEL FOR EVALUATION RATING BASED ON EVALUATION BUDGET	31
FIGURE 10. MEDIAN OVERALL RATING PER COMPONENT	31
FIGURE 11. EVOLUTION OF AVERAGE GLOBAL RATING PER COMPONENT AND TYPE OF EVALUATOR.....	32
FIGURE 12. AVAILABILITY OF KEY INFORMATION ON EVALUATION REPORTS (N=88).....	43
FIGURE 13. AVAILABILITY OF KEY INFORMATION ON TOR (N=54)	44
FIGURE 14. LEVEL OF AGREEMENT OF EVALUATION MANAGERS WITH STATEMENTS ABOUT EVALUATION PROCESS	48
FIGURE 15. KEY CHALLENGES PERCEIVED BY SURVEYED EVALUATORS	49
FIGURE 16. PROPORTION OF STAKEHOLDER FREQUENTLY OR ALWAYS USING KEY EVALUATION DOCUMENTS	52
FIGURE 17. PERCEIVED LEVEL OF READINESS OF STAKEHOLDERS TO CONDUCT EVALUATIONS.....	53

LIST OF TABLES

TABLE 1. INTENDED USERS AND EXPECTED USE FOR THE META-EVALUATION	17
TABLE 2. KEY AREAS OF ANALYSIS OF THE META-EVALUATION	18
TABLE 3. METHODOLOGICAL APPROACH AND STRATEGIES	188
TABLE 4. META-EVALUATION QUESTIONS AND TECHNIQUES.....	19
TABLE 5. DIMENSIONS UNDER ANALYSIS THROUGH THE META-EVALUATION TOOL	20
TABLE 6. SIX-POINT SCALE CONSIDERED FOR DIMENSION 1 ON QUALITY OF EVALUATION REPORTS	21
TABLE 7. SECONDARY DATA SOURCES (DOCUMENTS AND INFORMATION COLLECTED)	23
TABLE 8. KEY LIMITATIONS AND MITIGATION STRATEGIES	24
TABLE 9. OVERALL SCORING LEVELS AND LEVELS OF ACHIEVEMENT.....	25
TABLE 10. EVOLUTION OF GLOBAL RATING PER THEMATIC AREA.....	29
TABLE 11. OVERALL RESULTS FOR COMPONENT 1: EXECUTIVE SUMMARY.....	33
TABLE 12. OVERALL RESULTS FOR COMPONENT 2: PROJECT BACKGROUND.....	34
TABLE 13. OVERALL RESULTS OF COMPONENT 3: EVALUATION BACKGROUND.....	35
TABLE 14. OVERALL RESULTS OF COMPONENT 4: EVALUATION CRITERIA AND QUESTIONS.....	35
TABLE 15. OVERALL RESULTS OF COMPONENT 5: EVALUATION METHODS	36
TABLE 16. OVERALL RESULTS OF COMPONENT 6: EVALUATION FINDINGS	38
TABLE 17. OVERALL RESULTS OF COMPONENT 7: CONCLUSIONS.....	40
TABLE 18. OVERALL RESULTS OF COMPONENT 8: RECOMMENDATIONS	40
TABLE 19. OVERALL RESULTS OF COMPONENT 9: LESSONS LEARNED	41
TABLE 20. OVERALL RESULTS OF COMPONENT 10: GOOD PRACTICES	42

ACRONYMS

AVRR	Assisted Voluntary Return and Reintegration
CO	Country Office
DAC	Development Assistance Committee
EU	European Union
GMDAC	Global Migration Data Analysis Centre
GP	Good practice
HQ	Headquarter
ILO	International Labour Organization
IOM	International Organization for Migration
KII	Key Informants Interview
LL	Lessons Learned
M&E	Monitoring and Evaluation
OECD	Organisation for Economic Co-operation and Development
OIG	Office of the Inspector General
PPR	Project Performance Review
PRIMA	Project Information and Management Application
RBA	Rights Based Approach
RO	Regional Office
ROMEIO	Regional Monitoring and Evaluation Officer
ToR	Terms of Reference
UN	United Nations
UNEG	United Nations Evaluation Group

1. EXECUTIVE SUMMARY

META EVALUATION BACKGROUND

This report documents the results of the meta-evaluation of the International Organization for Migration's evaluations (IOM) during 2017-2019 (from now on, "the meta-evaluation"). The meta-evaluation was conducted between November 2019 and April 2020 by an external consulting firm (Artival Research & Evaluation).

In line with the ToR, the main objectives of the meta-evaluation are:

1. To assess the quality and use of internal and external evaluations (centralized and decentralized) conducted between 2017 and 2019, with the aim of providing actionable recommendations to enhance the quality and utilization of evaluations.
2. To provide IOM staff and partners with feedback on the quality of its evaluations, and if capacity-building efforts to support the conduct of evaluations have made a change in its quality.
3. To produce a full-fledged meta-evaluation tool in which to base the evolution of IOM's evaluation quality in the mid and long term.

The key questions of the meta-evaluations made reference to three main areas, (1) evaluation report quality, (2) evaluation process and capacities and (3) evaluation dissemination and use. The list of questions is included in the inception report, attached as annex 13. The meta-evaluation assessed all internal and external evaluation reports (centralized and decentralized) conducted for/by the IOM in the period 2017 – 2019.

The main users of this evaluation are the OIG Evaluation staff and the Regional Monitoring and Evaluation Officers (ROMEOS), whom together will be able to strengthen support and guidance in line with the findings of the report. In addition, this evaluation will be relevant for IOM's senior management as a mean of assessing how the Organization should improve the quality and use of its evaluation reports. Finally, IOM staff in general, and specifically programme managers and thematic specialists should take specific look at the quality and use of evaluations hence, this evaluation provides better understanding of this aspect and how this can be improved.

METHODOLOGY

The meta-evaluation follows a utilization focused approach¹ and was developed through a mixed-method concurrent triangulation² research strategy. The consulting team collected both quantitative and qualitative data concurrently and compared databases to establish convergences, divergences or a combination of both. Therefore, a process of confirmation, disconfirmation, cross-validation and corroboration was carried out in order to obtain key findings in which to base conclusions and recommendations. The three complementary

¹ https://www.betterevaluation.org/en/plan/approach/utilization_focused_evaluation

² Creswell, J. (2003) "Research design", third edition.

research techniques considered were the assessment with the scoring tool specifically designed for the meta-evaluation, the semi-structured interviews and the online survey launched.

A total of 88 evaluation reports were included in the assessment. With regard to the key informants for the interviews, the sample follows a purposeful sample approach, selecting those evaluation managers at the regional level and other relevant staff suggested by OIG representatives. Finally, the survey targeted every manager involved in evaluations during the period considered (2017-2019), as well as internal and external evaluators. The cross-cutting issues considered for the assessment are gender, rights-based approach, and environmental sensitivity and sustainability. The assessment determines the extent to which these issues are taken into consideration when designing evaluations, as well as when defining findings, conclusion and recommendations.

The main limitation faced was the limited number of evaluations available with information on budget assigned and training received by internal evaluator. This prevented from fully addressing question 4 of the assessment about the influence of budget on evaluation quality.

KEY FINDINGS

When it comes to the quality of evaluations, one third of the evaluation reports assessed meet or exceed quality requirements considered. Roughly half of the evaluation reports meet certain requirements but not all of them, and around 20% of the reports did not meet the minimum requirements of quality established. Moreover, there is a positive evolution over time in the quality of evaluation reports. The analysis of the evolution of component rating over time shows that the rating for all of the components increased for internal evaluations during the period under assessment. In addition, there are no relevant differences on the quality of internal and external evaluation reports. However, internally trained evaluators reach an overall rating considerably higher than the global average.

The components with the highest median score are “conclusions”, followed by “findings”, “criteria and questions”, “evaluation background” and “project background”. The components with the lowest median score are “lessons learned” and “good practices”, followed by “recommendations” and “evaluation methodology”. The great majority of reports, especially internal evaluation reports, follow the structure of the evaluation report template included in the IOM guidelines and the IOM Development Fund templates.³ Similarly, the Terms of Reference also use the IOM templates. The majority of reports include a gender equality approach, less than half of reports include Rights Based Approach (RBA) issues and mostly all fail to include environmental concerns. ToR do not usually make specific reference to cross-cutting issues, especially RBA and environmental issues.

Concerning the evaluation process, the decision on whether an evaluation should be internal or external does not follow a standardized procedure. In this line, there is no systematic approach in place to define in which evaluations ROMEOs should participate more, it is mainly a demand driven approach. The global opportunities identified by the analysis that help to

³ The key guidance provided for evaluation related templates is from the IOM Project Handbook (2017) however, the IOM Development Fund guidelines for evaluation also provide a variation of the evaluation template that can be used.

strengthening IOM's evaluation quality are the strong institutional allegiance and the rising network of internal evaluators. On the other hand, the main factors affecting the evaluation process are the limited funding which thus limits the timeframe available to carry out evaluations, together with the incipient evaluation culture within the organization. Concerning specific processes, the selection of evaluation criteria and methodology was reported very effective and fit for purpose, as well as the communication flows with relevant stakeholders during the evaluation. On the other hand, key specific challenges during the implementation period include mainstreaming cross-cutting issues, the limited availability of M&E data, the diversity of internal evaluation skill levels, the difficulty in accessing key informants, and the staff turnover. In general, the feedback process is considered essential to ensure quality and it is only negatively affected by the limited evaluation capacities of participants in some cases and lack of a screening process for receiving comments.

Regarding evaluation capacities, as the evaluation training was intended for internal evaluators, it has already reached a significant number of staff. Conversely, a considerable percentage of evaluation managers have not received any specific training in evaluation at IOM yet. The vast majority of IOM staff who have been trained are satisfied with the applicability of the training to their daily work. In this line, the most consulted evaluation related documents are the IOM Project Handbook and the IOM Evaluation Guidelines. The fact that other key documents such as the Guidance for Addressing Gender in Evaluations are not consistently used might explain some of the results of the quality assessment. Both evaluation managers and internal evaluators expressed high motivation and commitment to conduct evaluations. Nonetheless, managers perceived themselves less skilled than internal evaluators in evaluation.

Finally, when it comes to evaluation dissemination and use, the analysis reveals that the dissemination of evaluation results has been enhanced recently by introducing some new tools such as the evaluation briefs and the evaluation repository. The key factors identified in facilitating evaluation use are the institutional efforts made in recent years toward promoting actionability of evaluation recommendations, as well as the new PRIMA for all system which was only rolled out in 2019.⁴ On the other hand, key factors hindering evaluation use are the limited use of a tracking mechanism for existing recommendations, the reduced engagement of stakeholders during the evaluation for ownership and the scarce resources available for evaluations. Furthermore, additional management and institutional commitment in the promotion and use of evaluation results is required.

CONCLUSIONS

Each conclusion specifically addressed a meta-evaluation questions, and they have been numbered following the order of these questions.

Conclusion 1. When it comes to meeting UNEG standards⁵, the majority of **evaluation reports** satisfactorily presented what was evaluated and why (purpose and scope), how the evaluation was designed and conducted (evaluation questions, methodology and limitations), what was

⁴ Currently, there are no evaluations included in PRIMA.

⁵ As included in the document: United Nations Evaluation Group (2016). Norms and Standards for Evaluation. New York: UNEG 2017

found and on what evidence (findings) and what was concluded from the findings in relation to main evaluation questions addressed. On the other hand, the majority of reports fell short in adequately addressing cross-cutting issues throughout the process, including comprehensive recommendations and including what, if anything, could be learned from the evaluations (lessons learned). Evaluation report templates have positively evolved over the years.

Generally speaking, the great majority of **ToR** available meet most of the UNEG standards. However, management arrangements are not consistently included in the majority of ToR, neither (and more importantly) are specific concerns on cross-cutting issues, particularly environmental issues and RBA. The shortfalls found for evaluation reports and ToR might be explained by the fact that the evaluation reports and ToR templates included in the IOM Project Handbook do not address adequately these issues.

Conclusion 2. During the **evaluation design phase**, the use of templates for drafting ToR and inception reports is widespread, which is very positive to ensure consistency. One of the key limitations found was the lack of a systematic approach to define whether an evaluation should be internal or external, as well as a standardized criterion to determine which evaluations are mandatory to be followed up by ROMEOS, who usually work on a demand-driven basis.

As for the **implementation of the evaluation**, the overall key opportunities found are the strong institutional support toward the evaluation function and the richness of the emerging internal network of evaluators, which ensures that generated evaluation knowledge and skills built remain within the Organization. Concerning major challenges, the limited funding available for evaluations, together with the short timeframes usually established and an incipient evaluation culture with limited understanding of the potential uses of evaluations within the Organization were considered key elements that affect negatively the quality of evaluations.

Conclusion 3. The great majority of reports, especially internal evaluation reports, followed the structure of the template evaluation report included in the IOM Project Handbook of 2017. Similar is the case for the ToR. This had its advantages and disadvantages: following the templates ensured alignment to some of the key issues considered in the UNEG standards. Conversely, the reliance on these templates led to some of the shortfalls identified, with the consequence that those documents did not meet UNEG standards (see conclusion 1). Furthermore, it should be noted that **IOM guidelines include very useful templates beyond ToR and evaluation report templates that could increase the quality of the final evaluation products**. For example, IOM Project handbook include the evaluation matrix template, that is not consistently used and that has been proven to boost report quality. In addition, IOM guidelines include a comprehensive list of key questions suggested for incorporating cross-cutting themes into the evaluations, which has been hardly used in practice.

Conclusion 4. As included in the limitations section, it was not possible to analyse the influence of time and funds earmarked to evaluations and the way they influence evaluation quality. **Despite the fact that no imbalances in quality ratings were found between internal and external evaluations, and that higher evaluation budgets did not impact the quality of evaluation reports, the analysis is limited and no robust conclusions could be drawn.**

For future reference, when considering the budget of internal evaluations, the time

dedicated by internal evaluators should be factored in to establish the real cost associated to any given evaluation.

Conclusion 5. The greatest opportunities for improvement lie in the adequate inclusion of cross-cutting (CC) issues at all levels, including specific evaluation questions and methodologies, addressing these issues in key findings, and more importantly, reflecting this analysis in the final set of recommendations. Similarly, recommendations are also a key component on which to focus, not only by ensuring CC issues are taken into considerations, but also clearly establishing their actionability (what, how and when) and responsibility (who is called upon to act). The systematic inclusion of a specific section on lessons learned and good practices is also considered essential.

Conclusion 6. There is a positive evolution over time in the quality of evaluation reports. The analysis of the evolution of component rating over time shows that the rating for all of the components increased for internal evaluations during the period under assessment. The main reasons for this positive evolution are the introduction of ROMEOS, the institutional efforts made toward strengthening the evaluation function and the introduction of updated policies, manuals, guidelines and templates and the introduction of the evaluation training package in late 2017.

Conclusion 7. Generally speaking, there are no relevant differences on the quality of internal and external evaluation reports. Despite these similarities, evaluation reports conducted internally have experience a considerable positive evolution during the period 2017- 2019, something that was not identified for external evaluations. Additionally, **internal trained evaluators reach an overall rating considerably higher than the global average** for both internal and external evaluators, which leads to conclude that efforts made in terms of capacity building appear to be paying off. To complement this, surveyed staff expressed their satisfaction with the applicability of the training to their daily work, as well as their motivation and commitment. When it comes to gaps in evaluation competencies, evaluation managers have a lower perception than internal evaluators on their evaluation skills.

Conclusion 8. There have been recent efforts at IOM that will help to effectively **disseminate evaluation results internally**, such as the introduction of the evaluation briefs (2018), and the repository of evaluations for IOM staff (end of 2019) and will be further improved with the use of PRIMA for all database where evaluations will be stored. Concerning external stakeholders, all evaluations reports are publicly listed on the IOM evaluation website as per the evaluation policy, and all the evaluation reports will be accessible through the new website. Despite this, all groups of stakeholders consulted agreed on the need to keep strengthening the evaluation dissemination process, as well as promoting clearer guidelines and procedures for disseminating results. Concerning the **use of evaluations**, there is also a need to increase the internal communication and information knowledge management to improve the performance of the use and dissemination of evaluation results. Only by ensuring that more efforts are done to increase the engagement of management, related staff (chiefs of mission, regional directors and regional thematic experts) and other stakeholders, will have direct results in the accountability of the implementation of recommendations.

GOOD PRACTICES

The analysis reveals that a great majority of evaluation reports, and more specifically, internal reports, follow the templates included IOM Project Handbook and the IOM Development Fund guidelines. This reveals a **positive tendency on following the existing key documents** as a main framework when conducting evaluations. Despite the fact that some of these templates are not consistently used, it is expected that changes in those templates that are followed would bring about positive changes in evaluation quality.

The PRIMA for the IOM Development Fund system was developed as an initiative to deploy an institutional project information management solution to assist IOM to develop, manage, monitor, and report on Development Fund-funded projects and their results. More importantly for this assessment, all IOM Development Fund related evaluations are uploaded on the PRIMA for IOM Development Fund system, so it is available to anyone with access to the system around the world. Despite the fact that the system is not a repository and does not provide direct access to evaluation reports, it made evaluations accessible, and eventually led to the development of the new PRIMA for all system. This system, which has only been rolled out since 2019, is expected to contribute to the dissemination of evaluations results by linking evaluation reports to project documents, which could be very useful when designing and planning interventions.

The coaching initiative is part of the evaluation capacity building strategy launched by the Organization in recent years. Under this initiative, any internal evaluator can request the support and guidance of key experts and staff within the organization during the evaluation process. Although the first intention was to support internal evaluators on their first assignment, the initiative also gives the opportunity to any internal evaluator to request support to colleagues with the goal of increasing evaluation quality. This good practice amplifies the communication channels among IOM evaluation staff around the world and helps disseminating knowledge and debates, while, consequently, fostering staff and organization cohesion. In addition, the coaching initiative has a considerable potential not only to strengthen the capacities of staff, but also to develop evaluation culture.

LESSONS LEARNED

The scope of the meta-evaluation went beyond assessing quality of evaluation reports (and ToR), and also put focus on evaluation processes and use. This has given a broader scope of analysis that helped understand and explain some of the key constraints and limitations identified as a result of the quality assessment conducted. The exercise did not only serve the purpose of controlling the quality of evaluations, but also the assessment of their adequacy and opportunity to IOM context and their contribution to the achievement of the Organization's evaluation function. It provided a broader analysis and therefore it helped sustain and explain some of the findings with regard to the context and reality in which evaluations are conducted, ensuring a higher robustness for the whole analysis. All this should be considered for future meta-evaluation exercises, which should not only focus on analysing the evolution of evaluation report quality over time but also how processes and uses have changed compared to what has been presented in this meta-evaluation.

RECOMMENDATIONS

Recommendation 1: IOM should **update the inception, evaluation reports and ToR templates, and use them under all frameworks and donors.** The use of evaluation templates should remain mandatory for internal evaluators. As for external, it should be mandatory to include all specific sections considered in aforementioned templates. **The following should be considered:**

- **A standardized format to draft recommendations** should be added to the relevant section of the evaluation report template. The format presented in this recommendations section could be followed, as it includes the key elements considered toward ensuring the actionability and comprehensiveness of recommendations.
- **Include a specific section on lessons learned and good practices in evaluation report templates,** to ensure these are included in every evaluation report when applicable. Even when not applicable, the information on the lack of lessons learned and good practices would be also relevant to managers and other stakeholders. The International Labour Organization (ILO) templates for these items could serve as an inspiration⁶.
- Ensure that **all evaluations include an evaluation matrix** as a mandatory deliverable, and that reference is made to the relevant existing template in the IOM Project Handbook.
- Ensure **higher comprehensiveness** of evaluation reports and ToR by including evaluation costs, duration of the evaluation, a list of tables and figures and an attached inception report in evaluation report templates. ToR templates should include a specific section on norms and standards, specifically request the identification of unintended effects and include management arrangements.
- Use the **meta-evaluation tool designed as a checklist** to ensure that all key aspects (items) are taken into consideration for all kind of evaluations. This will help increase the ratings dramatically in the next global meta-evaluation, as the tool should serve as a basis for the assessment of future evaluation reports.

Recommendation 2: **Evaluations would provide meaningful information about cross-cutting issues if gender-responsive actions, RBA and environmental sensitivity would be consistently included in all the phases of the evaluation.** This should be considered from the early identification of stakeholders and definition of the evaluation methodology, to the final analysis of results and more importantly, as part of the recommendations. In order to ensure that, it is proposed to:

- Include an **additional criterion on cross-cutting issues in the ToR template,** as well as in the template for evaluations questions. A link or reference to the template on CC thematic evaluation questions already included in the IOM Project Handbook should be also made for reference.
- Lesson learned and good practices templates should offer ideas and experiences on how to detect and reports CC themes.

Recommendation 3: In order to complement already mentioned institutional efforts made in recent years, the following strategies should be considered toward strengthening the evaluation process:

- Establish a **systematic approach to decide when to conduct internal or external evaluations.** The type and timing of an evaluation could depend on the amount of funding,

⁶ https://www.ilo.org/global/docs/WCMS_165981/lang--en/index.htm

thematic considerations, the purpose or the duration of the project under consideration, and should be negotiated with the donor. Criteria used by other UN agencies (i.e. UNDP, UNICEF) could be used as a reference.

- Keep promoting evaluative approaches and PPR for **evaluations with reduced budget, and in an overall sense, ensure that evaluations with low budgets focus on very specific areas or information needs.** Complex, multi-country or regional evaluations should ensure a minimum evaluation budget allowing sufficient time for its adequate implementation. Among the strategies to maximize the efficient use of evaluation resources would be to conduct joint thematic evaluations covering similar interventions under same or different donors and combine evaluation budgets.
- **Keep implementing evaluation capacity building activities and the coaching initiative,** enriching the existing network of internal evaluators and increasing evaluation skills of general staff as stated in the OIG M&E strategy 2018-2020. This will also contribute to develop a strong evaluation culture within the Organization and limit the negative impact of staff turnover.
- **Establish clear guidelines to homogenize the involvement of ROMEOS in different evaluations** (complexity, duration of the project, budget available, etc.), ensuring that their involvement is not only based on a demand-driven approach, but also according to some shared criteria. This is particularly relevant when considering key evaluations for which ROMEOS in-depth involvement should be compulsory.
- **Keep developing and expanding a comprehensive and shared pool of evaluation experts,** both internal and external, including key thematic and geographical expertise, as well as working languages.
- **Recognizing the role played by internal evaluators.** The use of public recognition feeds, peer recognition and rewards could be complemented with other positive reinforcements, such as economic incentives, internal promotions, or allocating staff time to accomplish their evaluative function during key evaluation stages.

Recommendation 4: IOM would enhance evaluation dissemination and use by:

- **Providing access to a** systematic classification of evaluation results in the evaluation repository, allowing the identification of key findings, conclusions, recommendations, lessons learned and good practices by thematic and geographical areas, type of evaluation, etc. would be key to boost internal evaluation use for decision making and programming.
- **Implementing strategies to ensure the inclusion and participation of the target population in the evaluation process** (e.g. include them in ToR, allocate mission time for their participation, engage IOM staff in the inclusiveness of this target group). This could explore the relevance and strategic fit of the intervention outside the sphere of influence of donor/client/implementers, and make evaluations more according to RBA.
- **Continue raising awareness of the utility and uses of evaluation** among staff, as another strategy to build evaluation culture within the Organization. More specifically, efforts should be made to engage management, such as chiefs of mission, regional directors and regional thematic experts.

2. BACKGROUND

Introduction

This report presents the findings, conclusions and recommendations of the meta-evaluation of all evaluations conducted for the International organization for Migrations (IOM) from January 2017 to December 2019. The firm Artival Research & Evaluation carried out the consultancy between November 2019 and April 2020. The IOM reviewed the products, provided comments and signed-off the final report.

About the object of the meta-evaluation

As stated in the ToR, evaluation is an important learning and accountability tool on which there is an increasing emphasis placed both within the IOM and externally by donors and key stakeholders, encouraging the use of evaluation findings and recommendations in formulating policies and programmes. To provide more quality assurance system, efforts are being made to continue building a culture of evaluative understanding and practice, particularly given the increased acknowledgment of the importance of evaluations for reporting on project results, lessons learning and identifying innovative and replicable practices. These efforts made by IOM are:

- Since late 2016, eight out of nine IOM regional offices have Regional Monitoring and Evaluation (M&E) Officers who provide guidance, capacity building and technical assistance to country offices in their respective regions;
- The Office of the Inspector General's Central Evaluation (OIG/Evaluation) approved the OIG Strategy for the Management of its Evaluation and Monitoring Functions 2018 – 2020;
- The OIG Evaluation team has grown to continue providing additional support to and continuity across the regions;
- In December 2017, OIG launched the IOM internal evaluator training aimed at contributing to the creation and development of a pool of internal evaluators able to carry out evaluations. A second training was held in June 2018.

3. OBJECTIVES AND SCOPE

Objectives

The main objectives of the meta-evaluation are:

1. To assess the quality and use of internal and external evaluations (centralized and decentralized) conducted between 2017 and 2019, with the aim of providing actionable recommendations to enhance the quality and utilization of evaluations.
2. To provide IOM staff and partners with feedback on the quality of its evaluations, and if capacity-building efforts to support the conduct of evaluations have made a change in its quality.
3. To produce a full-fledged meta-evaluation tool in which to base the evolution of IOM's evaluation quality in the mid and long term.

The main users of this evaluation are the OIG Evaluation staff and the Regional M&E Officers (ROMEOS), whom together will be able to strengthen support and guidance in line with the findings of the report. In addition, this evaluation is relevant to IOM's senior management as a mean of assessing how the Organization should improve the quality and use of its evaluation reports. Finally, IOM staff in general, and specifically programme managers and thematic specialists should take specific look at the quality and use of evaluations, since the results of the analysis provide a better understanding of this aspect and how this can be improved.

Scope

In line with the ToR, the meta-evaluation assessed the information contained in the internal and external evaluation reports (centralized and decentralized) for the period 2017 – 2019 and the ToR (when available). This information was complemented with the feedback provided by key stakeholders through interviews and an online survey. A total of 88 internal and external evaluation reports were assessed. The geographic coverage includes all the different regions and sub-regions where IOM operates. More information about the evaluation reports assessed can be found in section 3 and figure 2.

The **meta-evaluation focused on maximizing the utility of its results and their actual use.**

In order to do so, the meta-evaluation design and process was carried out with careful consideration on meeting the information needs of primary and secondary audiences as tentatively expressed in table 1. In the interest of ensuring the optimal use of the findings, the meta-evaluation results are presented constructively and appreciatively to the possible extent. This is key since the meta-evaluation standards set for the exercise go beyond the evaluation requirements in place during the period under evaluation. This is justified by the fact that these standards should become the basis for future meta-evaluation exercises, allowing evaluation quality comparisons over time.

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

Type of User	Designation	Intended use
PRIMARY	OIG Evaluation staff Regional M&E Officers	<ul style="list-style-type: none"> ▪ Development of IOM's evaluation function. ▪ Improve preparation, implementation and resource management. ▪ Accountability ▪ Learning
	IOM's senior management and IOM in general	<ul style="list-style-type: none"> ▪ 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	<ul style="list-style-type: none"> ▪ Organizational learning/development
	Indirect beneficiaries / civil society	<ul style="list-style-type: none"> ▪ Learning/development, accountability

4. APPROACH AND METHODOLOGY

The purpose of this section is to establish the approach and methodology that guided the data collection, analysis, interpretation and the way in which the final deliverables were framed.

Approach

The driving questions for the meta-evaluation can be summarized as “What is the quality of the evaluations conducted and what is required to strengthen IOM’s evaluation function in the future?” and “How useful were the evaluations conducted for the IOM in the proposed period?” The following table reveals the key concerns under consideration 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	<ul style="list-style-type: none"> • Evaluation and ToR quality based on reference documents, and including cross-cutting issues • 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 on the quality of evaluation through time
Evaluation capacities	<ul style="list-style-type: none"> • Quality per type of evaluator
Evaluation use	<ul style="list-style-type: none"> • Strategies in place toward maximizing evaluation use. • Issues that facilitated or hindered evaluation use

Methodology

The meta-evaluation followed a utilization focused approach⁷ and was developed through a mixed-method concurrent triangulation⁸ research strategy. The consulting team collected both quantitative and qualitative data concurrently and then compared databases to establish convergences, divergences or a combination of both. Therefore, a process of confirmation, disconfirmation, cross-validation and corroboration was carried out in order to obtain key findings in which to base conclusions and recommendations. In practice, this means that quantitative data provided findings to trace statistical trends, whilst qualitative data obtained during the assessment process provided explanations and nuances of the quantitative trends found. The strategy was conducted in an iterative way to ensure the highest methodological robustness.

The following methodological strategies and data gathering techniques were used:

Table 3. Methodological approach

Methodological strategies	Research Techniques	Data gathering tools
<ul style="list-style-type: none"> ▪ Qualitative strategies (Discourse analysis) 	<ul style="list-style-type: none"> ▪ Semi-structured interviews ▪ Evaluators comments and perceptions 	<ul style="list-style-type: none"> ▪ Interview protocols ▪ Scoring tool
<ul style="list-style-type: none"> ▪ Quantitative strategies (Descriptive and inferential statistical analysis) 	<ul style="list-style-type: none"> ▪ Analytical framework ▪ Online survey 	<ul style="list-style-type: none"> ▪ Questionnaires ▪ Scoring tool

The methodological proposal solved to the extent possible methodological challenges by:

⁷ https://www.betterevaluation.org/en/plan/approach/utilization_focused_evaluation

⁸ Creswell, J. (2003) “Research design”, third edition.

- **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, and data gathered through qualitative methods (interviews, document review and the scoring tool) 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 used different sources of information in order to fill the gap between the absence of information in some cases, non-updated information or unreliable information. Data obtained from the assessment of reports and ToR was complemented by individual opinions (structured interviews and surveys). Different sources allowed triangulation of information especially when identifying good practices and recommendations for the future.

The three complementary research techniques considered were:

1. ***Assessment with the scoring tool:*** The analytical framework and its scoring tool provided the quantitative data in which to base the meta-evaluation results, allowing the emergence of statistical trends by means of proportions and measures of central tendency. Once the quantitative database was ready, specific software was used to conduct the analysis and produce the visuals (Tableau, Excel and Illustrator). The meta-evaluation tool also gathered qualitative information about perceptions of the evaluators that went beyond the rating given. This was reflected in the comments made to each particular rating.
2. ***Semi-structured interview:*** interviews with a select number of IOM staff was the other crucial method in the meta-evaluation, as it allowed the triangulation of information, point of views and opinions of different stakeholders. The views and information provided deepened the findings of the evaluation. A total of 17 interviews were conducted, targeting ROMEOs as well as a selection of key informants suggested by OIG and the reference group.
3. ***Survey:*** an on-line survey was developed and launched to complement qualitative information. Three different questionnaires were developed: for internal evaluators, for external evaluators and for IOM evaluation managers. The questionnaires were sent to a total of 22 internal evaluators, 71 evaluation managers and 22 external evaluators. In general terms, the response rate was high, obtaining a total of 97 complete questionnaires: 21 from internal evaluators, 57 from IOM evaluation managers and 20 from external evaluators.

The KII protocol and survey questionnaires are attached for reference. These are informed by the key meta-evaluation questions considered, as well as by the key areas under assessment included in the Scoring tool.

Table 4. Meta-evaluation questions and techniques

Meta-evaluation questions	Concept	Research Tools
1. To what extent are IOM evaluation reports and their ToR meeting the UNEG evaluation standards using evaluation criteria and including cross-cutting issues?	Quality of evaluations	Scoring tool Interview Survey

Meta-evaluation questions	Concept	Research Tools
2. On which quality aspects or factors do the evaluation processes (from preparation to reporting) and evaluation reports excel and where are they falling short?	Evaluation process	Scoring tool Interview Survey
3. Are evaluation reports and TORs developed based on the recommended templates provided by IOM guidelines?	Quality of evaluations	Scoring tool
4. Have the time and cost of/resources assigned to the evaluations influenced the quality of the final product?	Quality of evaluations	Scoring tool Interviews
5. What can be determined about the overall quality of the evaluation reports and where do the greatest opportunities for improvement lie?	Quality of evaluations	Scoring tool Interview Survey
6. During the period covered, are there any changes in the quality of the evaluations conducted and if yes, what factors may have contributed towards these changes?	Quality of evaluations	Scoring tool Interviews
7. What patterns of quality exist (if any) between those conducted by the trained internal evaluators, non-trained internal evaluators, and external evaluators? What strengths and gaps in evaluation competencies do the processes and reports indicate?	Evaluation capacities	Scoring tool Interview Survey
8. How were/are evaluations used and promoted? What factors facilitated or hindered the use of the evaluations?	Use of evaluations	Interview Survey

As revealed in table 6, the scoring tool specifically designed for this exercise played a major role in the analysis. The tool is based on the evaluation questions considered and a series of rubrics and ratings specifically developed to address and judge the extent to which each particular concept/evaluation question is addressed.

The analytical framework that served as the key tool to conduct the meta-evaluation considers four different levels of analysis:

- **Dimensions:** the first level of analysis includes four categories as described in table 5.
- **Components:** second level of analysis, there are the key sections in which the assessment of each dimension is based.
- **Items:** represent the general criteria consider to adequately assess the compliance with each of the different components; they indicate key concepts to be considered under each specific component.
- **Elements:** represent the ultimate sub-criteria in which the assessment is based. One item can comprise one or more elements, and the assessment is based on the compliance with the set of elements consider for each item.

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

DIMENSION	UTILITY
1. Quality of evaluation reports	<ul style="list-style-type: none"> ▪ Collect information about the quality of evaluation reports. ▪ Rate the quality and relevance of the information within reports.
2. Comprehensiveness of evaluation report	<ul style="list-style-type: none"> ▪ 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 proposed by IOM to evaluators.

Dimension 1 is the most comprehensive and provides useful information for the analysis on ten different components: the executive summary, project background, evaluation background, questions and criteria, evaluation methods, findings, conclusions, recommendations, lessons learned and emerging good practices (see annex 1 for more details). Each of the main components comprises one or more elements or criteria in which to base the assessment.

Once all elements under a particular component are considered, a series of rubrics were used to determine the level of compliance with established criteria. See next section for more details.

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

- A six-point scale included for Dimension 1 on quality of evaluation reports, which represents the main scoring system. The different rubrics are expressed in table 7. This is an ordinal scale and therefore the analysis would take into consideration frequencies, proportions, medians and average scores. Finally, a global rating was provided to each report considering the ratings given and the relative weight considered for each of the components (see figure 1).
- A dichotomous scale included for Dimension 2 and 3 (comprehensiveness of reports and ToR). This is a nominal scale and therefore the analysis considered mainly frequencies and proportions.

Table 6. 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 of the elements are present, and at least one of the elements present is incomplete	Somewhat unsatisfactory Not all of the elements are present, but those present are complete.	Somewhat satisfactory All of the elements are present but not all of them are complete	Satisfactory All of the elements are present and complete.	Highly satisfactory All elements are present, complete, interconnected and the report excels in covering the item.

In terms of enhancing the validity of the evaluation, the combination of proposed methods allowed the triangulation of information, strengthening the validity of the study. The evaluation team considered 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) was revised and presented for validation to key evaluation audiences.

In order to ensure the reliability of data-gathering tools, the evaluation team conducted several reliability procedures. The scoring tool and the different rubrics were piloted to test their

relevance and fit for purpose. All team members assessed the same evaluation report in order to:

1. Compare the scores given by each team member and the reasoned exposition of each component to ensure the identification of discrepancies in the interpretation of criteria and rubrics. This ensured the establishment of a common system of rating around the elements, items and components, and therefore the inter-rater reliability. Once all the interventions are reviewed, the extreme cases (evaluations with the highest or lowest ratings) were double-checked through a peer review process to ensure consistency.
2. Determine the level of effort required for the assessment of an evaluation to establish the final number of elements under analysis, and once confirmed by IOM, the final number of evaluation reports.

A series of rules were determined during the assessment to ensure inter-rater reliability. Those strategies are included in Annex 10. In addition, detailed information on the piloting process can be found in the inception report, attached as annex 13.

Figure 1- Relative weight for the rating of each component

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

Data sources

Primary data comprises the information produced through the analytical framework, KII and online survey designed (both quantitative and qualitative). Secondary data is considered all the information in which the rating and analysis is based, as described in table 7.

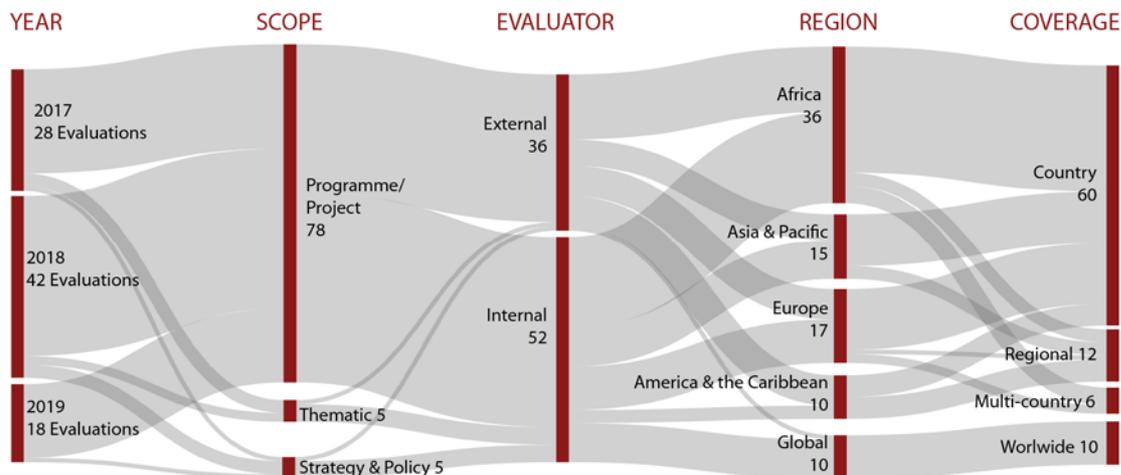
Table 7. Secondary data sources

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 (UN SWAP EPI, MOPAN, etc.) ⁹	Contextualization and deep understanding of key frameworks and strategies considered within IOM.
The evaluation reports and ToR	Key information in which to base the assessment and the meta evaluation

Sampling procedures

A total of 88 evaluation reports were included in the assessment. Some of the particularities of these reports (year, scope, etc.) are revealed in figure 2. The entire universe of evaluations conducted and shared with OIG during the assessment period is included. Therefore, no sampling strategies were needed. With regards to the key informant for the interviews, the sample followed a purposeful sample approach, selecting those ROMEOS at the regional level, OIG representatives and other key stakeholders that shared a vision at the global, regional and country levels. Finally, the survey targeted every evaluation manager involved in evaluations during the period considered (2017-2019), as well as internal and external evaluators.

Figure 2. Overview of evaluations under assessment



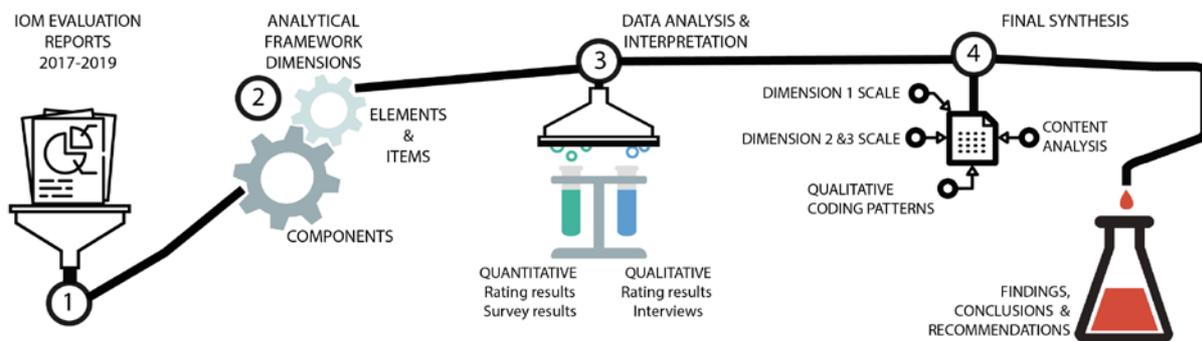
Cross-cutting issues

The cross-cutting issues considered for the assessment are gender, rights-based approach, and environmental sensitivity and sustainability. The assessment determined the extent to which these issues are taken into consideration when designing evaluations, as well as when defining findings, conclusion and recommendations.

⁹ The Multilateral Organisation Performance Assessment Network (MOPAN) assessed the performance of IOM 2017-2018 looking at IOM’s organisational effectiveness and results achieved against objectives. <http://www.mopanonline.org/assessments/iom2017-18/index.htm>

The following figure represents the **overall process for the assessment**, from the final universe of evaluation report under consideration to the findings, conclusions and recommendations of the final report.

Figure 3. Overall process of the meta-evaluation



Norms and standards

The consulting team ensured that the meta-evaluation process respects the UNEG basic premises of ethical and professional behaviour. The meta-evaluation was conducted following relevant IOM policies, as well as UNEG norms and standards¹⁰. The consultants carried out the assessment with integrity and honesty. The consultants declare not to have any conflicts of interest, as they have not directly or indirectly participated in any of the interventions under review, nor have a personal relationship with managers and/or consultants involved in those.

5. LIMITATIONS OF THE META-EVALUATION

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

Table 8. Key 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 by internal evaluator	The team included and analysis of the results of those evaluations with information on costs and trained/non-trained internal evaluators. The relevance of the analysis was based on the representativeness of the results.
Large range of migration thematic areas and sectors	The evaluations 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 took into account this challenge and different typologies to ensure homogeneity in the rating of the different elements. More on this on annex 10 on consistency issues.

¹⁰ <http://www.unevaluation.org/document/detail/1914>

6. FINDINGS ON EVALUATION REPORT QUALITY

Overall findings on evaluation report quality

This section presents the overall findings on the quality of the evaluation reports. The analysis on the quality of evaluation reports is based on the meta-evaluation tool specifically designed for the meta-evaluation. Some of the requirements included in the tool go beyond the requirements included in the UNEG checklist for evaluation reports and the templates included in the IOM Project Handbook. Therefore, the evaluations conducted in the period 2017-2019 did not have the reference quality checklist against which they have now been assessed. The OIG is fully aware of this circumstance and approved the approach with the idea that this first meta-evaluation exercise will establish the standards and quality baseline on which to base future analysis on the evolution of the Organization’s evaluation quality. To conclude, the consulting team presents the results of the quality assessment being fully aware of the aforementioned situation.

The global rating has been calculated considering the rubrics (a scale from 0 to 5) and the relative weights of each one of the components. The maximum rating possible to any report is 5, as shown in the following table.

Table 9. Overall scoring levels

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

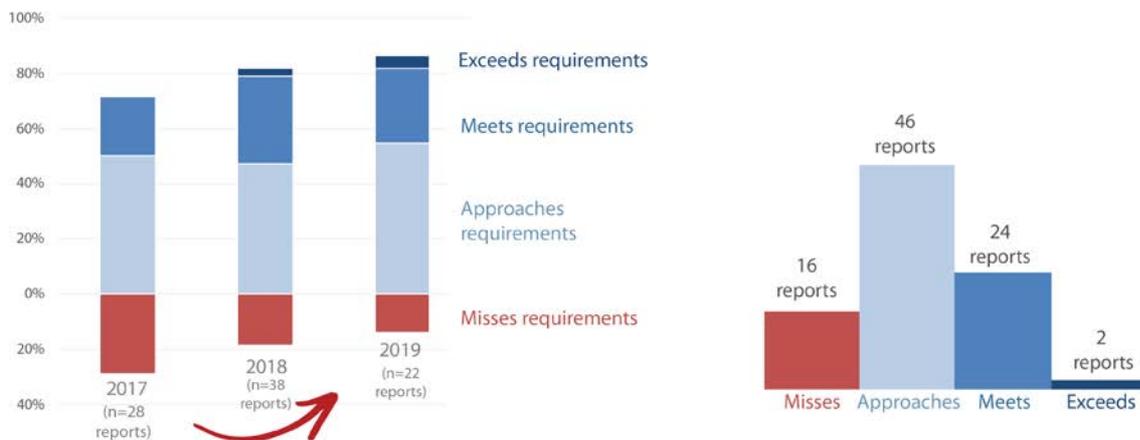
Key findings:

1. One third of the evaluation reports assessed meet or exceed quality requirements considered. Roughly half of the evaluation reports meet certain requirements but not all of them, and around 20% of the reports did not meet the minimum requirements of quality established.
2. There is a positive evolution in the quality of evaluation reports over time. The analysis of the evolution on the quality of the different components assessed shows that the rating for all of the components increased for internal evaluations during the period under assessment.
3. There are no relevant differences on the quality of internal and external evaluation reports.
4. Internal trained evaluators reach an overall rating higher than the average overall rating for internal and external evaluators.
5. The components with the highest median score are “conclusions”, followed by “findings”, “criteria and questions”, “evaluation background” and “project background”. The components with the lowest median score are “lessons learned” and “good practices”, followed by “recommendations” and “evaluation methodology”.
6. The great majority of reports, especially internal evaluation reports, follow the structure of the templates included in the IOM Project Handbook and IOM Development Fund template. These templates are often also use for ToR.

7. The majority of reports include a gender equality approach, less than half of reports cover RBA issues and almost all fail to include environmental concerns. ToR do not usually make specific reference to cross-cutting issues, especially RBA and environmental issues, however, gender is the most included approach.

One third of the evaluation reports assessed meet or exceed quality requirements considered. Roughly half of the evaluation reports meet certain requirements but not all of them, meaning they are of sufficient quality to be used but still have relevant limitations. Finally, around 20% of the reports did not meet the minimum requirements of quality considered for the assessment.

Figure 4. Evolution of global rating / Distribution of reports per rating over the three years.



The analysis of report quality over time shows a positive trend during the three-year period under consideration. A total of eight evaluation reports out of 28 did not meet the minimum requirements of quality in 2017, which represents 30% of the total. However, this percentage was reduced to 18% of reports in 2018 and to 14% of reports in 2019. In addition, one report in 2018 and one report in 2019 exceed quality requirements, something that did not occur in 2017. This positive trend is also identified in the overall median score per year, which increases from 2,4 points in 2017 to 2,75 points in 2019¹¹. Despite the positive evolution tracked, the aforementioned overall average rating of evaluation report in 2019 is still under the “approaches requirements” category, although close to reach the next category “meets requirements”.

The distribution of report per rating reveals that 80% of the reports fall within “approaches” and “meets requirements” level (see figure 4).

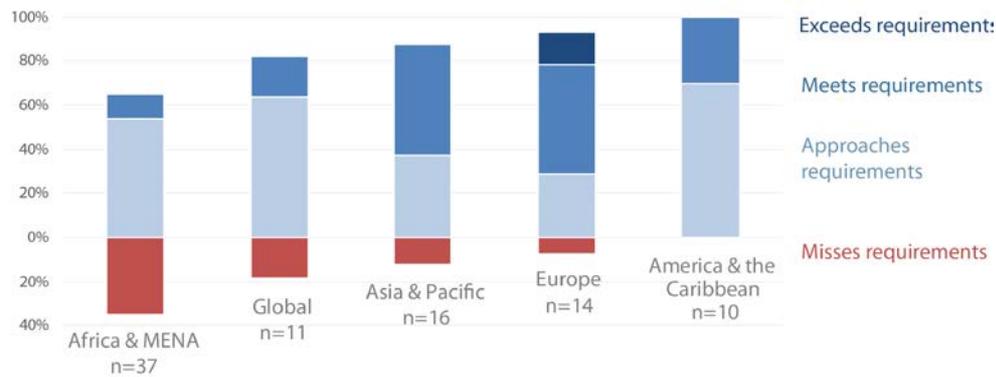
Geographical analysis

The distribution of results per region reveals that Europe and Asia & Pacific are the regions with a higher proportion of evaluation reports meeting requirements. In the case of Europe, is the only region with reports exceeding requirements (2 out of 14). America & the Caribbean is the only region with no report missing requirements, whereas global evaluations, and more

¹¹ The measure of dispersion considered when reporting medians is the inter-quartile range. The variability of results is low, as the inter-quartile range moves from 1 in 2017 to 0,7 in 2019.

substantively, evaluations conducted in the Africa and MENA region are the one having the highest proportion of evaluation reports missing requirements. Figure 5 reveal the aforementioned analysis.

Figure 5. Global rating levels per region



The median overall scores per region corroborate the analysis, as Europe and Asia and Pacific are the only regions with a median overall score that meets requirements (3,3 and 3 out of 5 respectively). The rest of regions reached an overall rating within the “approaching requirements” level, although America & the Caribbean region is very close to reach the next level. Finally, global evaluations and evaluations in the Africa & MENA region were the ones with the lowest overall median rating, although both remain in the “approaching requirements” level. It is important to notice that Africa & MENA was the region with the highest number of reports, which might explain to certain extent the variability of results.

When it comes to evolution over time, the overall median results per region during the years 2017-2019 does not show significant change in the case of America & the Caribbean and Europe. However, a positive pattern was identified in the case of Asia and Pacific, that moved from a median rating of 2,5 in 2017 to a rating of 2,9 in 2019 (see figure 7). In addition, Global evaluations moved from 2,2 in 2017 to a rating of 2,5 in 2019. A similar positive change was found in the case of Africa & MENA, with a median overall rating going from 2,2 in 2017 to 2,4 in 2019.

A more detailed analysis on the overall results per geographical regions can be found in Annex 12.

Figure 6. Global median ratings per region

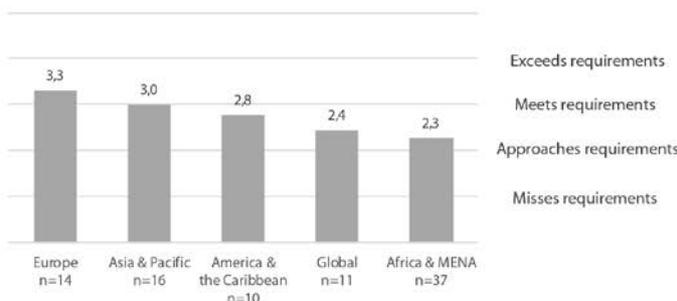
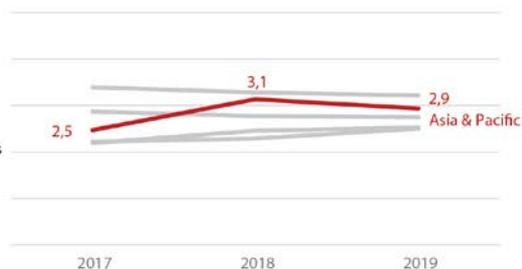


Figure 7. Evolution of median rating per region



Analysis by evaluator

The overall median score of internal and external evaluator is similar, and both groups share ratings within the “approaches requirements” level. Evaluations conducted by externals rated a median score of 2,6 and internal evaluations rated a median score of 2,8. In this line, the disaggregation of internal evaluations by OIG evaluations and non-OIG evaluations does not show any particular difference either, as both groups share similar ratings.

Evaluation reports carried out by internal evaluators who received training at IOM have an average rating higher than the global average¹². Reports of trained IOM staff have a median score of 3,1 (meeting the quality requirements considered for the assessment), which is 0,5 points higher than the overall median rating of all evaluation reports¹³. Nonetheless, the information on the number of internal evaluators trained is limited, as only 21 evaluations included that information. The evolution on the quality of reports by evaluator during the period 2017-2019 only shows a significant change for internal evaluations, whose overall median rating moved from 2 in 2017 to 2,9 in 2019, which means that their evaluations are very close to meet requirements.

When the geographical analysis of report quality over time is disaggregated by evaluator (internal/external evaluators), a relevant positive trend is illustrated in the case of internal evaluations conducted in Europe. The global median rating for this region moved from 3,1 in 2017 to 3,6 in 2019. Positive patterns in the quality of internal reports over time were also identified in the case of global evaluations (from 2 to 2,5 points over the aforementioned period).

Thematic analysis

Although most of interventions (and therefore evaluation reports) covered multiple thematic areas and sub-areas, each evaluation report was categorized by the evaluation team under one of the global thematic areas considered by the Organization. The global average results show **three thematic areas for which evaluation reports “meet quality requirements”, which are transition and recovery, migration and climate change, and migration health.** These three areas comprise a total of 13 evaluation reports (15% of the total). Nevertheless, it is important to point out that thematic areas with the highest median rating within the “approaches requirement” level (labour migration, immigration and border management and counter trafficking) are the ones comprising the highest number of evaluation reports (32 reports, 36% of the total). Conversely, the only thematic areas “missing quality requirements” (migration training and integration and migration research) are only represented by two evaluation reports, one for each category.

¹² Concerning the External evaluators, reports present a flat evolution on quality as they remain within the 2,5 range of the spectrum (approaching requirements) over time.

¹³ The standard deviation for the overall average rating is 0,62, which is low and represent little fluctuation on the ratings. As such, the mean is also an accurate representation of the ratings.

Figure 8. Global median rating per thematic area

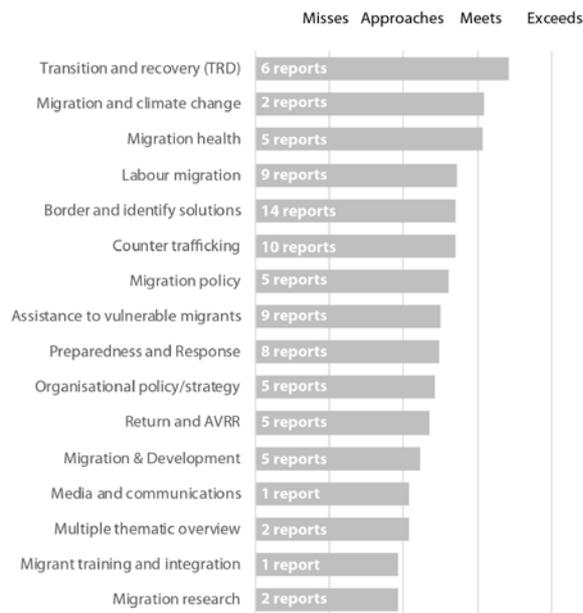


Table 10. Evolution of global rating per thematic area

THEMATIC AREA	2017	2018	2019
Transition and recovery (TRD)	3,3	3,2	3,7
Migration and climate change	2,9	3,2	n/a
Migration health	3,1	3,3	2,5
Labour migration	2,6	3,2	2,6
Immigration and Border Mgmt.	2,5	2,8	3,0
Counter trafficking	2,0	2,9	2,8
Migration policy	2,0	2,8	3,1
Assistance to vulnerable migrants	2,3	2,7	2,6
Preparedness and Response	2,3	2,5	n/a
Organisational policy/strategy	n/a	2,4	2,5
Return and AVRR	2,1	2,0	2,9
Migration & Development	2,7	2,2	1,4
Media and communications	n/a	n/a	2,1
Multiple thematic overview	2,1	n/a	n/a
Migrant training and integration	n/a	1,9	n/a
Migration research	1,9	n/a	n/a

Regarding the analysis on the evolution of quality over time, **relevant positive trends were identified for transition and recovery related evaluations, immigration and border management, counter trafficking, migration policy related evaluations and evaluations of interventions on return and AVRR.** The most significant changes over time shows an increase of 1,1 points in the rating of migration policy related reports during the period 2017-2019, followed by an increase of 0,7 point in the case of counter trafficking related reports. These two categories comprise a total of 15 evaluation reports. The positive evolution on rating involving the higher number of evaluations is for “immigration and border management” related interventions, that represent a total of 14 reports, whose overall average rating moved from 2,5 on 2017 to 3 in 2019.

The negative trends found on report quality over time involve migration health and migration & development related evaluations. The former experienced a reduction of 0,6 points during the period under consideration, whereas the latter underwent a decrease of 1,3 points. Despite this negative trend, the global median for migration health evaluations show that they meet quality requirements.

The analysis of the ratings disaggregated by donors was conducted between the IOM Development Fund interventions and all other interventions. A total of 89% of the reports (78 in total) could be classified under these two categories, (three reports were funded by both the Development Fund and the EU) and no information on donor was available in the rest of cases (10 reports). The results do not show major changes in the quality of evaluation reports, as **both the Development Fund and non-Development Fund evaluation reports obtained an overall average rating of 2,7** (within the “approaches requirements” level). The IOM Development Fund reports represent 35% of the total (31 reports out of 88).

Analysis per evaluation scope

In geographical terms, evaluation reports were categorized as global, regional, multi-country and country. The majority of evaluations were conducted at the country level (67%). There is an even proportion of evaluations covering the other three categories (11% of evaluations for regional, multi-country and global evaluations). The result of the analysis shows that all the categories fall within the range of 2,4 to 2,8 of overall median rating.

The analysis of report quality based on the geographic scope of evaluation shows no major changes over time for all the categories except one: multi-country interventions.

For multi-country evaluations, the overall median rating moved from 2,3 in 2017 (“approaches requirements”) to 3,3 in 2019 (meets requirements).

The timing of the evaluation was considered as an additional variable for the analysis, and five categories were introduced: mid-term evaluations, real time evaluations, final evaluations, ex-post evaluations and a category of “others”, mainly involving policy and strategic evaluations where timing of the intervention was not relevant. The majority of evaluations assessed (47%, 41 out of 88) are final evaluations. The second most common type of evaluation are ex-post (27%, 24 out of 88) and mid-term evaluations which represent 17% of the total (15 out of 88). Finally, evaluations under the “other” category represents 8% of the total (7 out of 88 reports), and only one real-time evaluation was assessed. The median overall ratings for all type of evaluations except those in the “others” group range from 2,6 to 3,2 points. Only evaluations under the “other” category are the ones with a global median rating close to “miss requirements”.

No major changes in quality over time can be found for the groups comprising the highest number of evaluations.

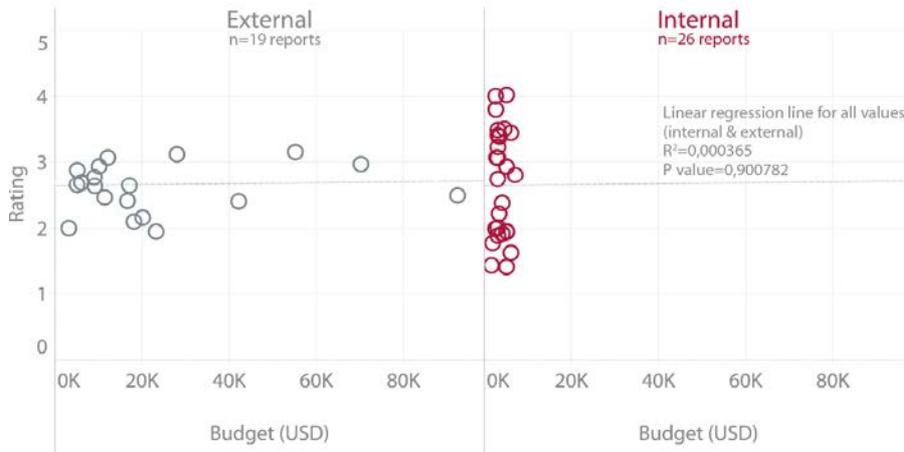
Finally, the majority of reports (89%) cover programmes and projects, and the rest cover strategies, policies and thematic areas evenly. While programmes and project related reports reach an overall average result of 2,7, this rating decreases to 2,4 for strategic & policy evaluations and to 2,2 for thematic evaluation reports.

Analysis per evaluation budget

The analysis of the linear relation between quality of the evaluation reports and the evaluation budget is one of the clear limitations of the Meta-evaluation. The information on evaluation budget was only available for 45 out of 88 of the evaluations assessed, which represents a clear limitation to determine robust conclusions. Nevertheless, the linear correlation model for the relationship between evaluation rating and quality is expressed in the following figure.

The figure divides the results by external and internal evaluation reports, although the trend line reflects the results of the analysis for all 45 reports.

Figure 9. Correlation model for evaluation rating based on evaluation budget



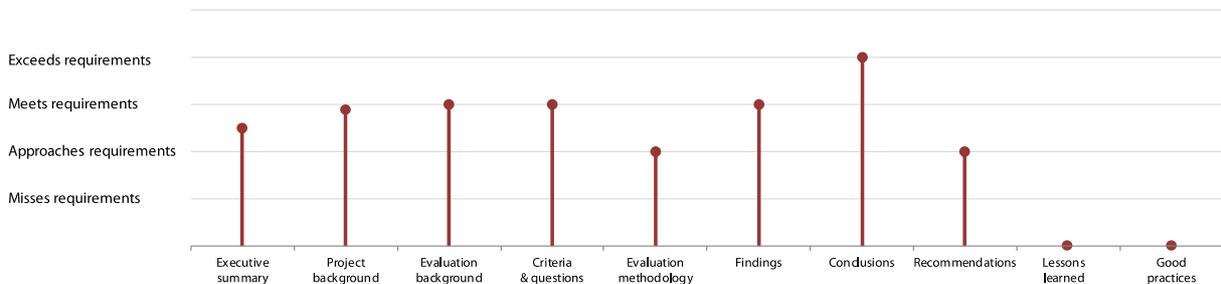
The figure shows a lower dispersion on the rating for external evaluation reports, whose scores range from 2 to 3,5. The scores for internal evaluation reports present a higher dispersion, ranging from 1,5 to 4. Conversely, external evaluations present a higher dispersion when it comes to budgets available, as they range from 3,000USD up to 92,000 USD. The budget available for internal evaluations remains within a narrower range, from 1,500 USD to 7,000USD.

Despite the aforementioned differences, both models present a linear correlation coefficient close to zero, meaning that changes in evaluation budget does not linearly correlate with changes in the rating of reports in the period 2017-2019. Nonetheless, the trend is not statistically significant¹⁴ at a 95% interval of confidence. Therefore, the conclusions for this analysis remain very limited.

Overall analysis per component

The overall median results per component under analysis reveals that **all components are rated within the “approaches requirements” level or above except in the case of lessons learned and good practices.** The element with the highest median rating is “conclusions”, followed by findings, criteria and questions, evaluation background and project background. The following figure shows data, and specific information about each of the components is provided in the following sections.

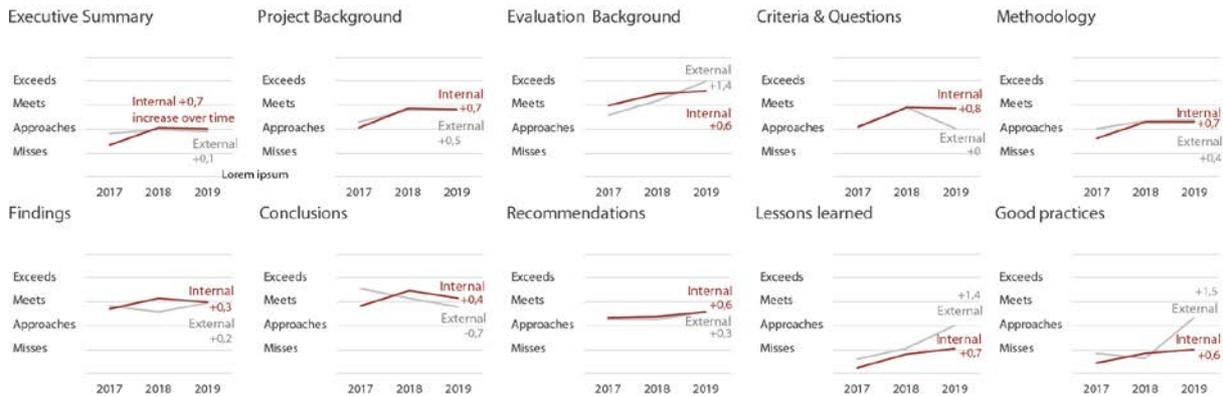
Figure 10. Median overall rating per component



¹⁴ Correlation considered significant at the 0,05 level (1-tailed), meaning P values <0,05.

The analysis of the evolution of component rating over time indicates that **scores of all the components increased for internal evaluations during the period under assessment**. This positive evolution is particularly accentuated when it comes to the elaboration of the executive summary, the project and evaluation background, evaluation criteria and questions, methodology, recommendations, lessons learned & good practices. The lowest rated components are those for lessons learned and good practises (within the “missing requirement” level) and the executive summary and evaluation methodology, both with a low rating within the “approach requirements” level; this was true of all evaluations, external and internal.

Figure 11. Evolution of average global rating per component and type of evaluator



Concerning external evaluations, ratings experience a positive trend over time in all components except when it comes to defining relevant evaluation criteria and questions (and including an evaluation matrix in the report) and conclusions, the latter with a considerable reduction of 0,7 points. On the other hand, there was a considerable positive increase on the rating over time of external evaluations in the quality of the evaluation background, lessons learned and good practices.

The following sections analyse each of the components in detail, based on the ratings given and the following scoring system:

HU	U	SU	SS	S	HS
Highly Unsatisfactory	Unsatisfactory	Somewhat Unsatisfactory	Somewhat Satisfactory	Satisfactory	Highly Satisfactory
(0)	(1)	(2)	(3)	(4)	(5)
NEGATIVE SPECTRUM			POSITIVE SPECTRUM		

Component 1: Executive Summary¹⁵

The component on executive summary comprises seven different items and represents 5% of relative weight of the total global rating. The table below show the different items and elements in which the assessment is based, highlighting the areas where the highest proportion of reports fall in a darker blue colour.

¹⁵ Annex 1,2,3 and 10 show both the tool and some clarification for consistency during the meta evaluation process.

Table 11. Overall results for Component 1: Executive summary

COMPONENT 1: EXECUTIVE SUMMARY – Weight 5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
ES.1 Explanation of the project background	1. Purpose 2. Objectives	26%	8%	9%	9%	38%	10%
ES.2 Overview of evaluation background	1. Purpose 2. Scope 3. Intended users	22%	19%	22%	23%	8%	7%
ES3. Concise description of evaluation methodology	1. Choice of methodology 2. Data sources 3. Limitations	30%	25%	36%	5%	3%	1%
ES4. Summary of all findings	All findings included in the report	25%	13%	13%	18%	19%	13%
ES5. Summary of all conclusions	All conclusions included in the report	36%	17%	6%	13%	14%	15%
ES6. Summary of all LL & GP	All LL and GP included in the report	81%	2%	8%	3%	3%	2%
ES7. Summary of recommendations	All recommendations included	32%	8%	3%	16%	10%	31%

Regarding the inclusion of the project background in the executive summary, the results are very divided between the two extreme scores: while a 26% of the reports do not include an explanation of the intervention background, 48% reach a satisfactory or highly satisfactory score. The rest of the reports are distributed among the remaining rubrics without exceeding 10% of the total (9 reports maximum per rubric).

Regarding the second item (overview of the evaluation background) results are spread mainly along the first four ratings (from highly unsatisfactory to somewhat satisfactory). Intended users of the evaluation are not systematically included in executive summaries. On the contrary, the scope of the exercise and the purpose of the intervention are almost always considered.

It should be underlined that most of the reports (91%) do not describe adequately the evaluation methodology in the executive summary. Moreover, one out of every three reports do not even include minimum information about this topic (highly unsatisfactory). Finally, six out of every ten executive summaries including information on methodology fail to provide key limitations faced during the process.

The summary of all evaluation findings and conclusions can be analysed as a whole. **Most executive summaries do not establish a clear division between findings and conclusions.** In most of the cases, findings and conclusions are described as one, and usually the information refers to conclusions rather than findings. In addition, the summary of findings and conclusions misses in some cases key information extracted from the report. The vast majority of executive summaries do not include a summary of lessons learned and good practices. This is mainly due to the fact that evaluation reports do not count with a specific section on these items in the report, and therefore these are not included in the executive summary.

Finally, with regards to evaluation recommendations, ratings are divided between the two extreme rubrics: 32% of reports are rated as highly unsatisfactory (which means recommendations were not included in the executive summary), while 31% are consider highly satisfactory (which means recommendations are included exactly as presented in the report).

Component 2: Project background

Divided in four items and nine elements, this component weights 5% of the total global rating. The component analyses the information that each evaluation report provides about the intervention under evaluation (evaluand). It must be noted that many reports do not include a project background section (about the project under evaluation) as an item itself, but normally as part of the evaluation background section. However, for the purpose of the meta-evaluation, this information was analysed separately.

Table 12. Overall results for component 2: Project background

COMPONENT 2: PROJECT BACKGROUND – Weight 5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
PB1. Brief outline of contextual factors	1. Socio-Economic context 2. Political context	30%	11%	8%	14%	27%	10%
PB2. Clear and relevant description of key stakeholders	1. Rightholders 2. Implementing agency (ies) 3. Duty bearers/Responsibility holders	13%	19%	10%	11%	39%	8%
PB3. Description of intervention logic	1. Objectives and results 2. Causal logic and assumptions	9%	11%	26%	19%	10%	24%
PB4. Funding arrangements	1. Reference to funding arrangements 2. Specific contributions of the IOM	31%	0%	8%	0%	52%	9%

A considerable number of reports (30%) do not include contextual information about the intervention under evaluation. On the other hand, reports rated within the positive spectrum normally merge socio-economic and political contexts in a brief paragraph. Regarding the description of key stakeholders, roughly half of evaluation reports fully present all key groups involved in or influenced by the evaluation.

One of the common pitfalls under this item is to include all stakeholders under the same label (usually “key stakeholders”) but without describing their roles in the intervention.

Even though reports usually present the objectives and results of the intervention under evaluation, **they rarely include a description of the causal logic and assumptions on how change is expected to emerge.** Most of the reports present results chains without further explanation, or even label those chains as “Theories of Change”, without including the logic and/or internal and external assumptions that enable results achievement.

Finally, the majority of reports (61%) include clear information of funding arrangement, including the budget provided for the intervention and donors. Some of these reports (8 in total) provide additional information and disaggregate budgets by results, actions, etc.

Component 3: Evaluation Background

The component about the evaluation background comprises three different items and represents 5% of relative weight of the total global rating.

Table 13. Overall results of component 3: Evaluation background

COMPONENT 3: EVALUATION BACKGROUND – Weight 5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
EB1. Explanation of purpose of evaluation	1. Purpose(s) 2. Explanation of evaluation purpose(s)	0%	5%	1%	9%	67%	18%
EB2. Description of evaluation scope	1. Geographic coverage 2. Timeframe 3. Thematic coverage	6%	2%	8%	45%	17%	22%
EB3. List of evaluation clients and main audiences of the report	1. Intended users (donors, partners, etc.) 2. Intended use per client	16%	10%	24%	16%	22%	13%

Most of the analysed reports adequately explained the evaluation purpose(s). In 85% of the reports, the information presented in this item is complete, and in another 9% the information is somewhat satisfactory, meaning the explanation of the evaluation purpose is briefly mentioned. Only 5 of the reports are negatively valued, mainly because the purpose of the evaluation is not adequately explained.

Regarding the evaluation scope¹⁶, the majority of reports are positively rated (84%). In 45% of the cases the information is incomplete due to not having explicit information on the thematic coverage, even if it is easy to infer from the context of the evaluation or the project. Information about the scope of the evaluation was not included at all in five cases.

Results are more diverse when analysing the list of evaluation users and audiences of the report. Half of the evaluations fall within the positive spectrum. Despite the fact that intended users are commonly included (84% of cases), they are usually only mentioned yet **information about the intended use per user is not included in roughly half of the reports**; noting that the use of an evaluation for a donor may be different than the use for an implementing partner. Nevertheless, an improvement over the years was identified in this regard, since the lowest-rated reports are from 2017 and 2018 (there are only two reports rated with 0 or 1 in 2019), while those in which the information is complete are from 2018 and 2019 (only five reports are rated as satisfactory or highly unsatisfactory in 2017).

Component 4: Evaluation criteria and questions

The component addressing evaluation criteria and questions includes three items and a total of six elements, as described in the following table. The component represents 7,5% of relative weight of the total global rating.

Table 14. Overall results of component 4: evaluation criteria and questions

COMPONENT 4: EVALUATION CRITERIA & QUESTIONS – Weight 7.5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
CQ1. Declaration of the evaluation criteria	1. Evaluation criteria 2. Justification of their use	5%	6%	6%	8%	44%	32%
CQ2. Relevance of evaluation questions	1. Questions address the goals and purpose of the evaluation 2. Questions include cross-cutting issues	28%	6%	15%	28%	7%	16%

¹⁶ The evaluation scope refers to clarifying the global coverage, the timeframe or thematic coverage of the evaluation.

CQ3. Inclusion of an evaluation matrix	1. Matrix includes indicators 2. Methods and sources of data per question	58%	3%	5%	5%	3%	26%
--	--	-----	----	----	----	----	-----

Roughly seven out of every ten reports reached the highest ratings (satisfactory and highly satisfactory) when it comes to presenting the evaluation criteria. This is mainly explained by the fact that including standard OECD/DAC criteria was considered sufficient, since the use of OECD criteria does not need to be justified as these are internationally standard evaluation criteria. However, when one of these criteria was not included, a clear justification was needed to be rated as satisfactory (e.g. in final evaluations the impact criterion is excluded or partially excluded from the evaluation).

Surprisingly, almost three out of every ten reports do not include specific evaluation questions, but rather mention the evaluation criteria. In some of these cases (7 out of 25 reports) there were no ToR attached to confirm that no questions were included. In other cases, evaluation questions were also missing in ToR. In addition, around 20% of the reports include questions about the three cross-cutting issues considered. Gender is the most common cross-cutting issues included in the evaluation questions, followed by RBA. Environmental issues are consistently neglected from evaluation questions, and only 2 reports include related evaluations questions.

Finally, a total of 30 evaluation reports (34%) included an evaluation matrix. It is interesting to note that **the average global score of evaluation reports including a satisfactory evaluation matrix is considerable higher than for those not considering this element** (from 2,4 global score to 3,1 global score).

Component 5: Evaluation Methodology

This component holds the highest number of items, eight in total, and represents 7,5% of weight of the total global rating.

The first item under consideration is the inclusion of an evaluation approach. **Almost half of the reports do not consider or mention any particular overarching evaluation approach** leading the exercise, such as utilization focused evaluation, theory led evaluations, case study, etc. Around 30% of the reports mention the type of approach considered, but the choice is not adequately justified. On some occasions, a certain approach is used because it appears specifically in ToR, with no further justification about its relevance. In other cases, the approach considered does not correspond with the methodology used.¹⁷ Only 17 reports (19% of the total) contain complete and adequately justified information about the evaluation approach.

Table 15. Overall results of component 5: evaluation methods

COMPONENT 5: EVALUATION METHODS – Weight 7,5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
EM1. Statement of the evaluation approach	1. Approach used 2. Justification	45%	13%	15%	8%	10%	9%

¹⁷ Rating in this section was based on the methodological description provided by the evaluator coupled with how the methodology was reflected in the evaluation findings, conclusions and recommendations. Hence, in some instances, although an evaluation may say it is using a specific approach, such as utilisation-focused, the evaluation fails to specify specific uses per user, a key component in which to base the approach.

EM2. Type of analysis considered	1. Type of analysis 2. Reliability assurance	10%	8%	16%	19%	32%	15%
EM3. Description of data sources	1. List of data collection methods and explanation 2. Description of different data sources	7%	2%	8%	18%	30%	35%
EM4. Sampling procedures	1. Sampling procedures 2. Justification for the selection	14%	7%	10%	32%	20%	17%
EM5. Inclusion of cross-cutting issues	1. Gender 2. Human Rights 3. Environmental issues	56%	20%	19%	3%	0%	1%
EM6. Stakeholder participation	1. Rationale for stakeholder participation 2. Level of participation	13%	48%	11%	9%	15%	5%
EM7. Limitations of the evaluation	1. Limitations 2. Mitigation strategies	23%	5%	17%	10%	30%	16%
EM8. Description of evaluation norms and standards	1. IOM & UNEG Norms & Standards 2. Data protection principles	72%	3%	8%	1%	9%	7%

Information on the type of analysis¹⁸ considered is satisfactorily addressed on 66% of the reports (all the elements are provided in 47% of the evaluations). Reports usually include a mixed method with a focus on qualitative strategies. However, **information on reliability assurance¹⁹ is usually incomplete or not explicitly stated**. In 19% of cases, information about the type of analysis is either not included or unsatisfactorily included; for instance, if an evaluation mentions reliability assurance however does not provide indication within the report to guarantee it.

Data sources are mentioned in the vast majority of the projects. 65% of the cases provide information in a satisfactory or very satisfactory manner. In 18% it is considered somewhat satisfactory since the description of the different data sources was incomplete or superficial, and in 8% no description is provided. Only 8 of the analysed reports (9% of the total) did not mention the data sources or did it unsatisfactorily. **The inclusion of an evaluation matrix always resulted in a positive rating on the description of data sources**, as it provides key information about stakeholders considered and techniques used to answer each evaluation question.

Sampling procedures are adequately included in 69% of the cases. The element most of the evaluations fail to comply with is the justification for the selection of a certain sampling strategy. **More than half of the reports do not include any specific methodological strategies to ensure the adequate inclusion and mainstreaming of cross-cutting issues.** In 20% of reports, cross-cutting issues are mentioned without establishing strategies toward their consideration throughout the evaluation process. In some cases, a gender responsive approach is used for the analysis (gender-disaggregated data) or collection of the information (specific disaggregation of male-female staff). The rights-based approach (or human rights as a cross-cutting theme) is mentioned in 15 evaluations, and the “do no harm” principle in two reports.

¹⁸ Type of analysis would include explaining the type of qualitative or quantitative data analysis, the statistical approach etc. within the evaluation.

¹⁹ Reliability assurance includes describing triangulation of data collection methods.

There is no mention of specific methodologies to address environmental issues in any of the analysed reports²⁰.

In some reports (48%) there is no explicit information about stakeholder participation: **the participation as sources of information is the only participatory approach considered**. Moreover, only 20% of the evaluation reports include a complete description about stakeholder's consultation process and participation in the evaluation.

Limitations of the evaluations are satisfactorily included in 56% of evaluation reports (in 46% the information is complete). Mitigation strategies are incomplete or missing in one third of the cases, and no mention to limitations was found in 20 projects.

The description and inclusion of evaluation norms and standards is explicitly addressed in 16% of the evaluation reports. Nonetheless, 72% of the reports do not have any information in that regard. Evaluation reports for which norms and standards are included in their ToR have a considerable higher rating under this item than those reports not including these aspects in their terms - 12 out of the 14 reports that adequately formulated norms and standards included specific relevant information in their ToR.

Component 6: Evaluation Findings

This component includes four items and is the one with the highest relative weight (25%).

Table 16. Overall results of component 6: evaluation findings

COMPONENT 6: EVALUATION FINDINGS – Weight 25%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
F1. Completeness	1. All evaluation criteria and questions are addressed 2. Findings aligned with purpose, questions & approach	0%	1%	15%	19%	30%	35%
F2. Robustness	1. Findings justified by evidence 2. Disaggregation of data by key variables	0%	5%	17%	20%	33%	25%
F3. Identification of causal factors	1. Causal factors leading to achievement or non-achievement of results are identified 2. Description of unintended effects	7%	3%	42%	19%	23%	6%
F4. Cross-cutting issues addressed	1. Gender 2. Rights-based approach 3. Environmental issues	22%	30%	25%	18%	5%	1%

The majority of evaluations (65%) present a set of findings that adequately address the evaluation questions and sub-questions. Most evaluations rated within the unsatisfactory spectrum included findings aligned with the purpose and approach of the evaluation, but did not address directly stated evaluation questions. The purpose(s) of an evaluation are mainly

²⁰ The IOM Project Handbook aims to contribute to the strengthening of IOM's capacity to, among others, "develop high-quality proposals that incorporate IOM's cross-cutting themes", but also recognise that "not all cross-cutting themes would be equally relevant for one project". The assessment gave major importance to cross-cutting issues as so IOM can promote coherent policies for sustainable results. Despite the fact that the relevance of cross-cutting issues might vary depending on the type of intervention, their omission should not be the norm. Thus, although the main recommendation is to try mainstreaming all cross-cutting themes, it would be necessary to include a very good justification for their absence.

related to accountability, improvement or knowledge generation. The evaluation questions are the specific set of information needs that help addressing those purposes. It is very possible to address to some extent the evaluation purpose without explicitly addressing key evaluations questions. For example, an evaluation question on value for money to assess efficiency (outcome level) may be addressed by identifying the level of budget expenditure (activity level). Examples of topics not covered in findings although specifically described in the evaluation questions are impact related questions (such as unintended effects) and efficiency related questions (such as value for money related questions).

Concerning the robustness of evaluation findings, the assessment shows that **findings are in general evidence-based and include certain level of data disaggregation**. In some reports, findings were considered not completely justified with evidence as they mainly or fully rely on the perceptions of informants, without considering additional pieces of evidence or data, or justifying the lack of data thereof. A total of 19 evaluation reports fell within the unsatisfactory spectrum mainly due to their lack of disaggregation of key data.

The third item of the component considers the extent to which findings identify causal factors leading to results. The ratings given are balanced between the positive and negative spectrum, as roughly half of reports fall within each category. The main reason this item is not fully complied with is the fact that unintended effects are not considered in 27 reports, and partially considered in 11 reports. Not all the ToR specifically request this piece of information, but it is considered a key issue that should be included in all evaluations. This was the reason to assess its inclusion irrespective of their specific inclusion in the Terms of Reference.

Evaluation findings do not adequately address all three cross-cutting issues in 94% of the evaluation reports. Moreover, 77% of the evaluations (67 reports) are rated within the unsatisfactory spectrum when it comes to mainstreaming cross-cutting issues.²¹ When analysed separately, the majority of reports included a gender equality approach (70% of evaluations) followed by 40% of reports including RBA issues and 6% of reports including environmental concerns. In any case, it is important to note that **when cross-cutting issues are included in the ToR, they are always considered in the evaluations**. However, the analysis shows that gender issues were not considered in 43% of the available ToR (25 out of 58 available documents). Similarly, the fact that environmental issues were only specifically included in 5 ToR out of the 58 available, might be linked to the limited consideration of this issue in evaluation reports (80 reports and their ToR did not consider this issue at all).

Finally, **addressing cross-cutting issues separately as another evaluation criterion has a considerable positive impact in the rating of related items**. However, this is not the common approach, as most of the evaluations mainstreamed cross-cutting issues into the standard OECD criteria.

Component 7: Conclusions

²¹ Please note that this rating is based on the expectation that all evaluations should address cross-cutting issues even if not mentioned in terms of reference.

The meta-evaluation tool considers two items to analyze the quality of the evaluation conclusion: the added value of the arguments presented, and the level of reasoning reached, with two elements each.

Table 17. Overall results of component 7: conclusions

COMPONENT 7: CONCLUSIONS – Weight 20%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
C1. Value	1. Conclusions are clearly linked to findings	6%	7%	17%	16%	38%	17%
	2. Provide insights to related findings						
C2. Reasoned	1. Conclusions reflect evaluative judgments	7%	3%	10%	25%	49%	6%
	2. Judgments are to the extent possible objective						

In general, **conclusions were mainly rated within the positive spectrum**; 62 reports positively rated when it comes to the value of the conclusions and 70 reports for the quality of the reasoning.

Data shows that **the majority of reports satisfactorily include conclusions linked to findings**, that provide and added value in relation to findings. Conversely, 11 out of 88 evaluation reports were rated as highly unsatisfactory under this item as no apparent or clear link was made between some of the findings and the conclusions. Additionally, 14 evaluation reports included brief conclusions that were not considered sufficient to provide added value to existing findings, as there were primarily summaries of the key findings and lacked additional interpretation and/or evaluative judgment. Regarding the extent to which conclusions are adequately reasoned, **80% of the reports have conclusions revealing critical thinking** through objective judgments. This means that conclusions are not only a summary of key findings, but also an interpretation of them that leads to a reasoned judgment in which recommendation are based.

Component 8: Recommendations

The component on recommendations comprises four different items and nine elements in which the rating is based, as described in the following table. The component represents 20% of weight of the total global rating.

Table 18. Overall results of component 8: recommendations

COMPONENT 8: RECOMMENDATIONS – Weight 20%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
R1. Clarity	1. Are clear and concise	0%	0%	6%	11%	69%	14%
	2. Are based on findings and/or conclusions						
R2. Relevance	1. Address key issues and are useful, tied to the object and purposes of the evaluation.	0%	3%	52%	15%	26%	3%
	2. Address cross-cutting issues						
R3. Responsibility	1. Specify who is called upon to act	22%	43%	28%	3%	3%	0%
	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	23%	11%	50%	13%	1%	2%
	2. Recommend a time frame						

The results of the analysis reveal that **the majority of evaluations include recommendations that are clear and concise, and that are clearly linked to the evaluation findings and conclusions**. Only five evaluations were considered within the negative spectrum of the rating (somewhat unsatisfactory) for this particular item.

Despite the great majority of evaluation recommendations address key issues and are tied to the evaluation purposes, the relevance of recommendations present a reduced overall quality reached. This is mainly due to the fact that **roughly 50% of the evaluations do not consider cross-cutting issues (gender, RBA or environmental issues) in their recommendations when relevant**.

Additionally, **the great majority of evaluations (84%) fail to adequately include recommendations that clearly establish responsibilities, means for achievement and levels of priority**. As described in the results for the item number three, 96% of evaluations (82 in total) are rated within the negative spectrum. Despite the majority of recommendations indicate somehow who is called upon act, they consistently lack general information on the means needed for the implementation (40% of the total) and are not prioritised (67% of the total, 59 evaluations).²²

Finally, when it comes to the actionability, **recommendations usually indicate the specific course of action need to be implemented (80%) but they lack a recommended timeframe (short, medium, long term) for their implementation (64% of the total)**.

This information is important as some of the aforementioned elements (timeframe, responsibility, priority) are expected to be included in the management response matrix which is filled out after the completion of the evaluation report. The management response matrix was not rated as part of this meta-evaluation however, only four of the reports assessed included such a document.

Component 9: Lessons Learned

Three items were included to analyse the inclusion and quality of the lessons learned in the evaluation reports. **Reference to lessons learned is absent in the majority of the evaluation reports**. A total of 33 evaluation reports (out of 88) included specific reference to lessons learned, ten of which did not meet the minimum requirements set. This is a relative low number considering that the majority of ToR specifically request lessons learned as a goal of the evaluation (in total, 30 out of the 54 ToR available). In the 23 reports where lessons learned were adequately included, the best rated element was relevance, due to the fact that they were derived from findings and represented relevant information.

Table 19. Overall results of component 9: lessons learned

COMPONENT 9: LESSONS LEARNED – Weight 2,5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
LL1. Relevance	1. Lessons learned are derived from findings 2. LL represent a relevant / new piece of information to be considered in the future.	64%	8%	3%	6%	17%	2%

²² Although this information does not have to be too detailed, it is considered important to mention in the recommendations.

LL2. Delimitation	1. Lessons capture the context from which they were derived 2. LL target specific users	70%	10%	5%	8%	7%	0%
LL3. Applicability	1. Lessons suggest what should be repeated/avoided in future contexts to guide action. 2. The LL include causal factors.	70%	8%	11%	3%	6%	1%

Concerning the delimitation of lessons learned, only nine reports captured both elements considered (the context from which they were derived and targeted specific users). With regards to the applicability of lessons learned, 10 reports suggested what should be repeated/avoided in the future or included causal factors leading to results.

This evidence shows that **very few reports included a high-quality section on lessons learned**; only two reports were scored with “highly satisfactory” for the relevance and applicability of their proposal.

Component 10: Good Practices

The component addressing good practices includes three items and a total of six elements in which the assessment is based, with a relative weight of 2,5%.

Table 20. Overall results of component 10: good practices

COMPONENT 10: GOOD PRACTICES – Weight 2,5%							
ITEMS	ELEMENTS	HU	U	SU	SS	S	HS
GP1. Delimitation	1. The good practices concisely capture the contexts from which they were derived 2. The good practices specify target users	64%	6%	10%	8%	11%	1%
GP2. Applicability & replicability	1. Description of how to implement the GP and by whom 2. The good practices explore the applicability in different contexts.	73%	5%	9%	9%	3%	1%
GP3. Impact	1. The good practice is linked to specific impacts 2. The different impacts identified are realistic	68%	8%	1%	10%	11%	1%

Generally speaking, good practices are not specifically addressed in evaluation reports.

No specific section is included or reference is made to good practices, as revealed in the results of the assessment: 65% of the reports do not include any information or reference in this regard, as reference to good practices could on found in 31 out of the 88 reports.

Good practices were to some extent included or adequately included in 17 evaluations reports. In these cases, **contextual factors were presented to explain how the good practices were identified, and did effectively identify target users.** This information is usually complemented with some indications of potential impacts the practice might generate. However, when it comes to applicability and replicability, only 12 reports included satisfactory information on this regard.

In any case, the evolution of results per year for related items experience a slight improvement, as it moved from the “highly unsatisfactory” area in 2017 to the “unsatisfactory” average in 2019 for all the three items under consideration. Good practices experienced a positive evolution over time (2017-2019) for both internal and external evaluations. However, the average score

for this item positively evolved from “missing requirements” to “approaches requirements” in the case of external evaluations, whereas the average score for internal evaluations remained within the “missing requirements” level.

Comprehensiveness of the Evaluation Reports

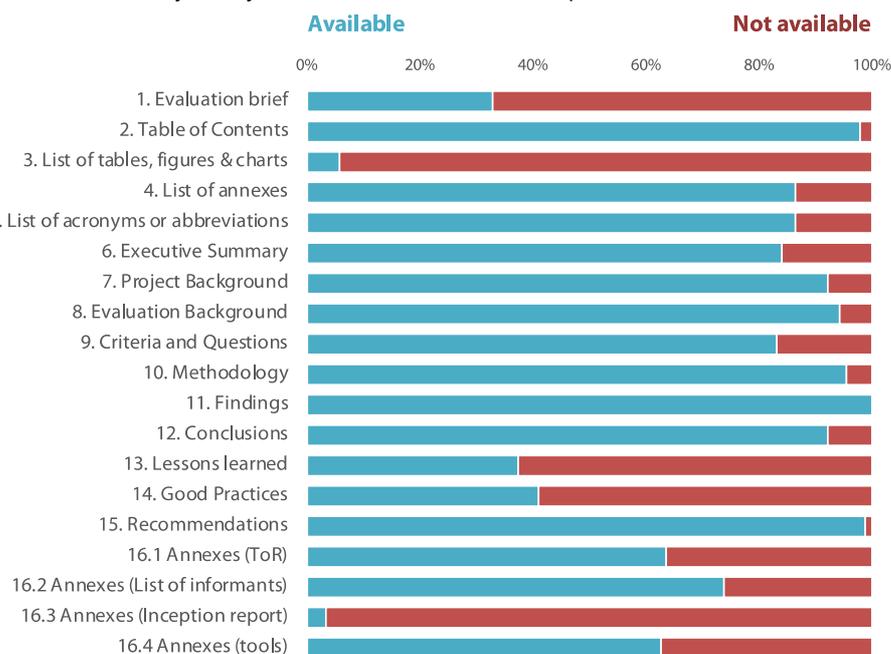
The comprehensiveness of the evaluation reports has been measured according to 19 elements. The evidence shows that **there are some gaps concerning the structure and presentation of the reports**. Evaluation briefs were introduced in late 2018 and are considered a key evaluation product. A total of 29 evaluation reports had an evaluation brief; 14 for 2018 and 2019, and only one in 2017. A list of tables and figures are not used in the vast majority of reports, and only six included them.

The key annexes considered (ToR, list of informants, inception reports and tools) are **partially included**. Around 60% of reports annexed ToR, list of informants and tools developed during the evaluation process. On the other hand, the inception report is practically never attached as annex. Only 27 reports included lessons learned and good practices, 16 of which included both elements.

On the other hand, **there are key elements that are present in the vast majority of reports (80%)**, namely the list of annexes, acronyms or abbreviations, the executive summary, the project and evaluation background, the evaluation criteria and questions, evaluation methodology and conclusions. Finally, the section on findings and recommendations were included in all reports except in one case, where no recommendations were included.

The figure below shows the proportion of evaluation reports that adequately included the different elements considered.

Figure 12. Availability of key information on evaluation reports (n=88)



Comprehensiveness of the Terms of Reference

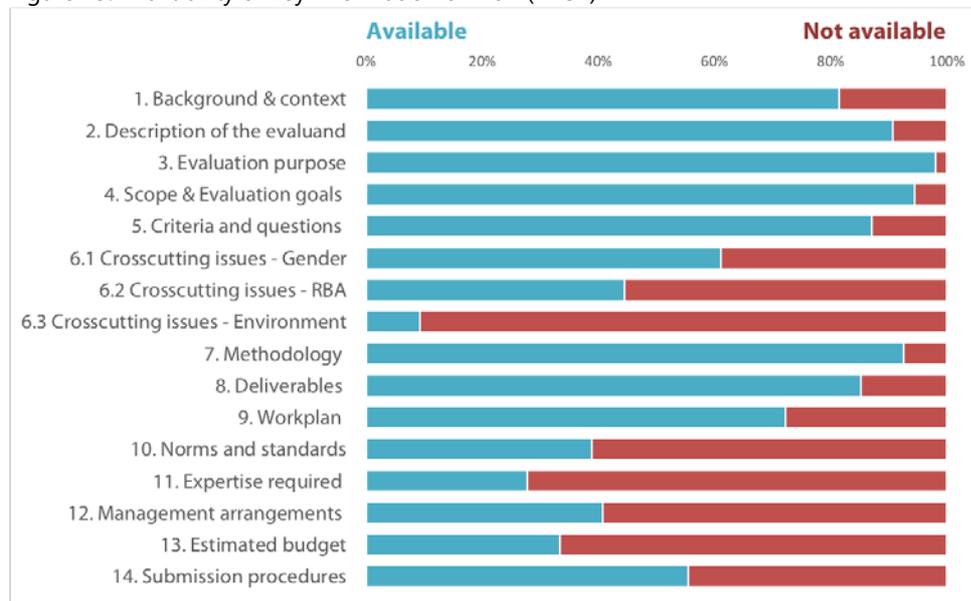
The analysis of the comprehensiveness of ToR is based on 14 different elements. The following figure shows the key trends on the availability of required information. It is important to notice that, out of the 88 evaluation reports analysed, 34 did not include ToR, even though in some of these reports ToR were listed as part of the annexes²³.

The information lacking in ToR is mainly specific questions and information on environmental issues. **In fact, the lack of specific reference to cross-cutting issues in ToR (either as evaluation criteria, questions, or others) is the key finding for this section**, as information on RBA and gender are also missing in 56% and 40% of available ToR respectively.

Expertise required to conduct the evaluation is only included in 15 out of the 54 available ToR, although it is understood that it might be not relevant in the case of internal evaluations; it should also be noted that the IOM Project Handbook recommends a separate call for applications where this information may be included²⁴. Moreover, ToR usually do not present an estimation of the overall evaluation budget, as only 33% of the ToR include it. As per the IOM Project Handbook, mentioning the budget is at the discretion of the evaluation manager²⁵. In some instances, only the fact that IOM would bear the cost of the evaluation is mentioned.

Finally, specific mention to evaluation norms and standards to follow during the evaluation and management arrangements are also usually not included, as these were included only in 38% of the ToR available.

Figure 13. Availability of key information on ToR (n=54)



It is interesting to note that the majority of ToR make specific reference to the need of identifying lessons learned and good practices as part of the evaluation. The analysis

²³ Important to note that this could be because some annexes were not available for the consultant team.

²⁴ IOM Project Handbook (2017), Module 6

²⁵ IOM Project Handbook (2017), Annex 6.1

reveals that 30 out of 54 ToR specifically request lessons learned, whereas 26 out of 54 ToR request the identification of good practices. As a good practice, some ToR specifically requested evaluation matrix as part of the annexes were found (2 cases).

7. FINDINGS ON THE EVALUATION PROCESS

The results presented in this section focus on the evaluation process, from design to reporting. Findings are mainly based on the analysis of the online survey and the key informant interviews, which are complemented with the desk review of key documents on the internal evaluation process.

Key findings:

8. There are no standardized procedures to determine when an evaluation should be internal or external. In addition, there is no systematic approach in place to define which evaluations ROMEOs should get involved in, and currently this involvement is demand-driven.
9. The global opportunities identified that strengthen IOM's evaluation quality are the strong institutional allegiance and the rising network of internal evaluators. On the other hand, the main factors negatively affecting the evaluation process are the limited funding which indirectly leads to limited timeframe in which to carry out evaluations, together with an incipient evaluation culture with limited understanding of the potential uses of evaluations within the organization.
10. Concerning specific processes, the selection of evaluation criteria and methodology in the ToR was considered by key informants as very effective and fit for purpose, as well as the communication flows with relevant stakeholders during the evaluation. On the other hand, key specific challenges during the implementation period include mainstreaming cross-cutting issues, the limited availability of M&E data, the existing internal limitations in evaluation skills, the difficulty in accessing key informants during the evaluation, and the high staff turnover.
11. In general, the feedback process on evaluation reports is considered essential to ensure quality and it is only negatively affected by the limited evaluation capacities of participants in some cases and lack of a screening process for receiving comments from the perspective of the evaluators.

Evaluation process: evaluation design

Since 2011, all IOM project proposals are expected to clarify if they intend to conduct an evaluation or not and justify if they do not. Central evaluations are selected based on their institutional importance, strategic considerations and risks associated. Central and corporate evaluations are managed and conducted by the OIG. Decentralized evaluations are managed by country, regional or other HQ units. When it comes to conducting decentralized evaluations, **the selection of the evaluator (internal or external) does not follow a standardized procedure.** Donor requirements is one of factors in deciding whether evaluations should be internal or external. The introduction of IOM Development Fund projects and the requirement to conduct ex-post evaluations under this framework increased the number of internal evaluations given the high number of projects and the limited budgets that these projects could afford for evaluation. In any case, the IDF does not follow a systematic approach to define when

to conduct internal and/or external evaluations. Some donors specifically allocate a M&E budget, while others do not consider evaluations when budgeting. When faced with evaluation budget constraints, internal evaluations are usually conducted, as they are considered less expensive. This is due to the fact that staff honoraria are not allocated under the evaluation budget, but only travel costs and daily allowances. Finally, the complexity of the intervention and the availability of external evaluators were also reported to be factors considered when deciding on the type of evaluation.

According to the results of the KII, the ToR are usually designed by either the evaluation manager or the programme manager (who can be the same person) or the donor. In all cases ROMEOs or OIG can provide their support. ToR are usually designed based on existing templates. Evaluation managers (whether they are programme managers or not) and Chiefs of Mission are responsible for the final approval of the choice of evaluators, with the support/validation in some cases of the ROMEOs, the OIG office and the donors.

The ROMEOs network was introduced in 2016, and their main role is to provide support in M&E issues to the different country offices, as well as evaluation quality assurance and coaching. They are acknowledged to provide key technical expertise and guidance when conducting evaluations, and they also conduct internal evaluations themselves. They design regional evaluation planning documents to track evaluations. However, ROMEOs do not have the capacity to be involved on all the evaluations carried out in a region, and **there is not a systematic approach to define to which extend ROMEOs should follow up on**. Currently the approach is demand driven, based on the requirements expressed by programme managers and/or evaluation managers.

Changes are usually not made to the ToR once the evaluator or evaluation team is selected. Nevertheless, a great proportion of evaluation managers (40%) stated that evaluation questions are usually added or modified during the design phase, as well as to a lesser extent data collection strategies and evaluation criteria; this may be done within the inception report or evaluation matrix.

Evaluation process: evaluation implementation

In order to analyse the **key overall opportunities and challenges affecting evaluation quality**, the survey and key informants addressed specific questions. The former included a question for the three groups surveyed (evaluation managers, internal and external evaluators), who were asked to express their level of agreement (six grades, from fully disagree to fully agree) with several statements about the evaluation process. The results reveal clear patterns among the three groups of informants, that were complemented by the key informants:

Key opportunities:

- A **strong institutional allegiance** toward evaluation generated over the past years, framed by the recent evaluation policy, the focus on Result based Management (RBM), the inclusion of specific evaluation roles within IOM's structure and a resolute commitment toward

developing evaluation skills and culture. The standardised support for M&E focuses on providing institutional guidance, ensure quality checking and enabling technical support.

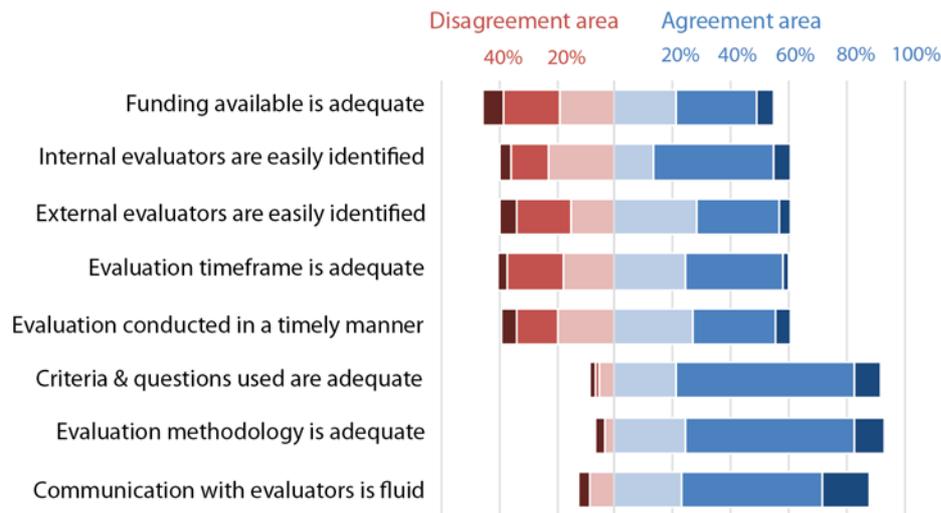
- **The internal network of evaluators** is perceived as a major asset, which ensures that generated evaluation knowledge and skills built remains within the Organization. Strong efforts have been made in recent years to build capacities and raise awareness on the importance of the evaluation function, as well as to create an enabling environment for strengthening IOM's evaluation function. More on this is included the following section on evaluation capacities and skills.

Key challenges:

- **The limited funding** allocated to conduct decentralized evaluations is considered by interviewees and surveyed informants one of the key constraints ensuring quality evaluations. The costs of evaluations are typically covered by a specific budget within the project, and evaluation allocations were in some cases reportedly reduced to limit the impact on the global intervention budget.
- As a consequence, key informants concurred that evaluations are conducted in a **short timeframe**, which limits the breadth and depth of the exercise, affecting the quality of the final product.
- **Incipient evaluation culture.** Informants expressed their satisfaction with the efforts made in recent years toward promoting a positive and constructive culture of evaluation in the Organization. However, this perception seems not to be dominant across missions, since evaluations are perceived in some cases as an unavoidable requirement for accountability, or in the worst-case scenarios, a ticking-the-box exercise that does not bring added value. In this line, several external and internal evaluators also pointed out the negative reaction evaluations generate due to limited openness to self-criticism. Conversely, monitoring is positively perceived as a way of ensuring project implementation, goal achievement and reporting.

The following figure shows the answers provided by evaluation managers in the online survey, which are very much aligned to those provided by internal and external evaluators.

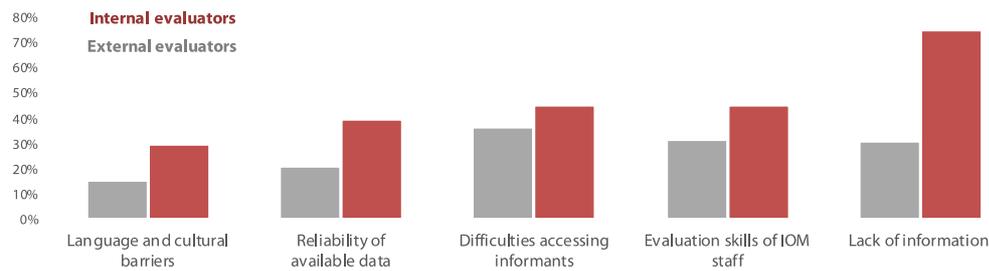
Figure 14. Level of agreement of surveyed evaluation managers with statements about evaluation process



When it comes to **specific evaluation processes that excel**, surveyed informants expressed their satisfaction with **(1)** the selection of the evaluation criteria and questions, combining OECD/DAC criteria with humanitarian related criteria and, in some cases, considering cross-cutting issues as an independent criterion. **(2)** The selection of evaluation methodology is also considered very relevant, as well as the fact that evaluators are allowed to modify or introduced new methodological strategies and techniques to complement the evaluation. Finally, **(3)** communication with evaluators (in the case of managers) and communication with IOM staff during the evaluation process (for internal and external evaluators) is also perceived as fluid throughout the process.

Concerning the **main challenges perceived during the implementation of the evaluation**, both internal and external evaluators shared similar concerns. **(1)** The lack of information about the intervention being evaluated (M&E data, secondary data, etc.) is considered by both groups as the major limitation when carrying out evaluations. This concern was also raised up by ROMEOs and other key stakeholders during the interviews: the availability of M&E data remains limited. The next constraint highlighted by surveyed evaluators and managers and corroborated through KII is **(2)** the diversity of evaluation skills of IOM staff, which leads to inconsistencies in the quality of evaluation reports. This is reflected in the high dispersion of the quality ratings for internal evaluations (see figure 9). As per KII results, only some country offices have specific expertise in evaluation. Informants also pointed out **(3)** the difficulty of accessing key informants. **(4)** Staff mobility and turnover was also shared as a key constraint faced toward ensuring quality evaluations, as key knowledge might be lost if adequate handover mechanisms are not in place.

Figure 15. Key challenges perceived by surveyed evaluators



One of the key processes on which the analysis focused is the **mainstreaming of cross-cutting issues**. All the three groups of informants shared the same pattern when ranking which cross-cutting issues are consistently considered during the evaluation process. Gender is considered the most common cross-cutting issue introduced in evaluations, followed by RBA and environmental concerns. This is clearly aligned to the results about the extent to which cross-cutting issues are considered in the findings section of evaluations and recommendations (see section 6.7 and 6.9 respectively). The feedback provided through the open-ended question was categorized and some key ideas about the reasons for not including cross-cutting issues in evaluations were identified, as follows:

- **Perceived relevance considered:** the idea that not all cross-cutting issues apply to all kinds of interventions is quite widespread among informants. However, the fact that not all cross-cutting issues would be equally relevant for one intervention does not mean they are to be omitted. Informants consider gender and RBA components relevant in all cases, but environmental issues are perceived as irrelevant in certain thematic areas. As expressed by an informant: "Gender and rights are fundamental elements of project design regardless of focus, whereas environmental elements are sometimes not associated within the project design".
- **Knowledge and ownership:** Additionally, evaluation managers and internal evaluators considered that despite recent efforts, cross-cutting issues (and particularly environmental sensitive and sustainability) are still not well understood within the Organization, and there is also a considerable lack of awareness. Time constraints when conducting evaluations is also considered a major limitation in not including cross-cutting issues adequately, especially when these are not perceived as a priority for the evaluation (as in most cases, environmental issues were considered).
- **Institutional guidance:** Environmental (and other cross-cutting issues) are also not adequately mainstreamed due to a lack of clear institutional guidance on how to approach them and apply them in practice. An example provided was the limited knowledge on what environmental-related indicators to apply in different contexts.
- **Overlooked.** Both internal and external evaluators also highlighted that cross-cutting issues were not considered since they were not in the ToR.

When it comes to providing feedback on the draft evaluation report, programme managers are the key stakeholders in the process. They are the ones who validate evaluation reports. Relevant IOM staff in country and regional offices are considered the second key group that provides comments to the evaluation reports, followed by ROMEOS, usually as per project managers requests. Finally, relevant staff at HQ (mainly OIG) and donors are also considered

key stakeholders to a certain extent, and invited to participate in the process. The Chief of Mission was also mentioned in some cases as a key stakeholder by both interviewees and surveyed staff. Feedback from national counterparts is mostly considered relevant for external evaluators (20% claimed reports are usually shared with them, compared to 5% of internal evaluators). However, sometimes only programme staff are given the opportunity to comment on the draft final report, and only once it is finalised, the report is sent to the rest of stakeholders.

Additionally, **more than 70% of evaluations manager and internal evaluators agreed or strongly agreed that comments received to drafts were valuable for improving the overall quality of evaluation reports.** This proportion is reduced to 60% in the case of external evaluators. Interestingly, around 5% of respondents from the three groups claimed not to have received comments to evaluation reports. The analysis of the open-ended question concerning the feedback process reveals the following shared ideas (in order of importance / number of mentions):

- All three key groups of informants agreed that the feedback process allows key informants to add additional information or data to the evaluation report, fill information gaps, enrich information and address factual errors. The whole process is considered essential to improving the overall quality of evaluations.
- Another shared opinion is that the feedback process enhances the engagement and ownership of participant stakeholders. It is considered key to contextualizing the evaluation results and validating their findings.
- Some evaluation managers considered that the feedback process helps prevent biased internal evaluations. They agreed that external stakeholders generally provide more objective inputs, while staff more closely associated potentially introduce bias.
- The limited evaluation capacities at the country office level was also brought up by internal and external evaluators as a key challenge to properly carrying out quality control measures.
- Some external evaluators claimed to have received a considerable number of comments, and that no focal point was appointed for screening and consolidating feedback received, which led to difficulties during the process.
- A minority of internal and external evaluators mentioned that no relevant comments or no comments were received at all.

8. FINDINGS ON EVALUATION CAPACITIES AND SKILLS

In recent years, the IOM has made an important effort toward improving the quality of evaluations. For this purpose, and among other strategies, several guidelines, handbooks and templates were designed and/or reviewed, and an internal training process was introduced to further develop the increasing network of internal evaluators and evaluation managers. The following section presents key findings on this particular area, as extracted from the results of KII and the online survey.

Key findings:

12. The internal evaluator training has put focus on training evaluators, and a majority of them have completed it. Conversely, a considerable percentage of evaluation managers have not received any specific evaluation training yet from IOM. Out of those who received training, the vast majority of both groups are satisfied with the applicability of the training to their daily work.
13. The most consulted evaluation related documents are the IOM Project Handbook and the IOM evaluation guidelines. The fact that other key documents such as the guidance for addressing gender in evaluations are not consistently used might explain some of the results of the quality assessment.
14. Both evaluation managers and internal evaluators expressed high motivation and commitment to conduct evaluations. Managers perceived themselves less skilled than internal evaluators in evaluation matters.

Internal evaluator training

The OIG M&E strategy 2018-2020 includes one key goal toward increasing staff, knowledge in and resources for M&E across the Organization. This subsection focuses on the efforts made towards increasing knowledge and more specifically, on the internal evaluator training that has been running since late 2017.

Proportionally speaking, evaluation managers have received less specific evaluation training than internal evaluators. More specifically, when it comes to internal evaluator training, survey results reveal that only 26% of evaluation managers have completed it (15 managers), compared to 70% of internal evaluators (14 in total). Conversely, evaluation managers (who are in most cases programme managers), have received in proportion more training on project development (86,67% of the project managers), project management (46,67%) and M&E e-learning course (40%) than internal evaluators²⁶.

During the interviews, several informants argued that there are IOM staff who works in evaluation with no specific knowledge about M&E and, though they appreciate the trust and responsibility given to them, that could also be stressful at times. In any case, they also recognize the important efforts IOM is putting to reach different staff and promote evaluation culture within the Organization.

Concerning the contents of the internal evaluator training, staff who completed the training expressed their satisfaction, as **90% consider it was useful or very useful toward improving their work on evaluation**. Only 3 people preferred not to or just did not answer this question. Nonetheless, they also highlighted that the training mainly focuses on technical and methodological aspects without mentioning specific considerations in the Organizations' thematic areas. To complement this idea, several interviewees pointed out the important role that thematic experts can play in the evaluation process. In addition, they suggested that the debate should focus on the reason and usefulness of the evaluations and not so much on their methodology, because often evaluations are just taken as an accountability tool.

²⁶ It must be noted that the Project Development and Project management trainings both have modules on evaluation, and the e-learning course on M&E has one module dedicated to evaluation managers.

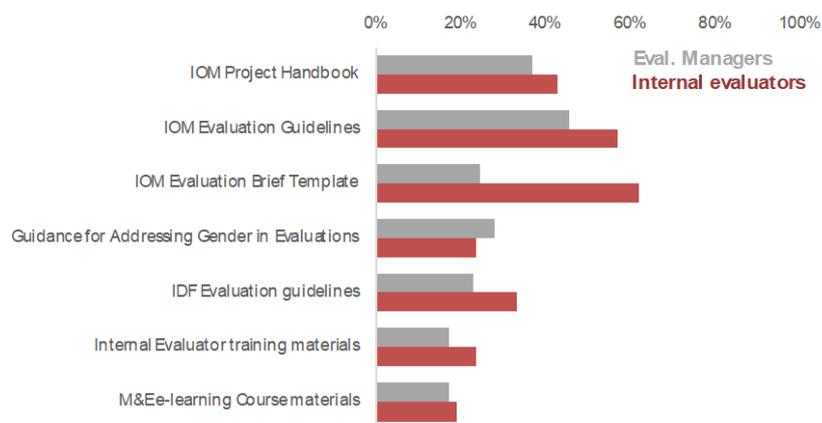
As explained by interviewed staff, **a good practice resulted from the training: the possibility of being coach by any IOM colleague during the evaluation process** (managers, evaluators, ROMEOS, thematic experts, M&E officers, etc). This coaching is mandatory for internal evaluators when conducting their first evaluation. More on this issue is included in section 11 about good practices.

Use of key evaluation documents

When it comes to using key evaluation documents, **the IOM Project Handbook and the IOM Evaluation guidelines are the most commonly consulted**. This is consistent with the results of the quality assessment, as most of the evaluations followed for example the evaluation report template included in the IOM 2017 Project Handbook. As expected, evaluation brief templates are used mainly by evaluators. A considerable number of evaluation managers recognised they never or rarely consult the other four key documents included in the questionnaire (guidance for addressing gender in evaluations, IDF evaluation guidelines, internal evaluator training material and M&E training materials²⁷). This is also consistent with the results of the quality assessment, especially when it comes to gender issues. It should be noted that almost half of the evaluation managers consulted never use the IDF Evaluation Guidelines. In any case, it is not a worrisome fact, since it is a three-page document with continuous references to the most widely used guidelines and handbooks. Interviewees corroborated these finding, and added the perception that a document is only made visible when its use is mandatory.

For internal evaluators data is more divided. A considerable percentage claim to always or often use IOM main guidelines, while it is less frequent for them to consult the training and e-learning materials. It should be mentioned that those who indicated they review other documents from the above mentioned, do consult UNEG evaluation guidelines and checklists. The following figure reveals the patterns identified.

Figure 16. Proportion of stakeholder frequently or always using key evaluation documents



External evaluators were also asked about evaluation documents that were shared with them during the evaluation process. Their answers are not very different from those given by internal

²⁷ The materials for the M&E e-learning course and the internal evaluator training are only available to those who have taken either trainings.

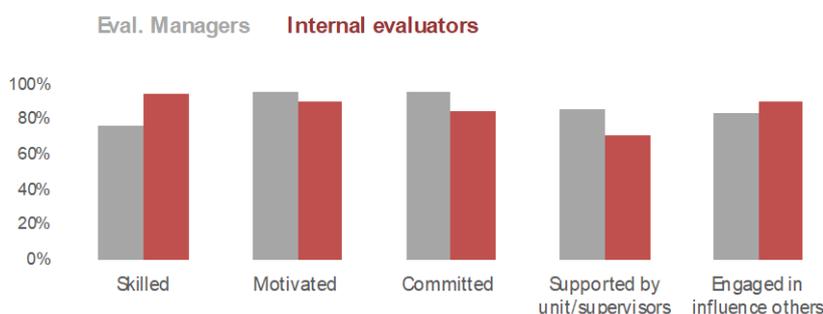
evaluators: most of them claimed to have used the IOM Project Handbook (42%) and IOM Evaluation Guidelines (37%). The evaluation brief template was shared with a 21% of them and the IDF Guidelines with 16%. As it was highlighted during the interviews, it is not mandatory for external evaluators to use these documents, though often suggested. The most striking piece of information for this section is that, **according to external evaluators, no one shared with them the Guidance for Addressing Gender in Evaluations.** This would not be worrisome if it were not for the results of the quality assessment, where it was found that only 4 evaluations had proposed methodologies that included cross-cutting issues.

Self-perceived preparedness of internal evaluators

This subsection analyses the self-perception of internal evaluators about their readiness to conduct evaluations. The section is based on the results of a specific question included in the survey, and for which respondents needed to rate their level of preparedness on five categories: skills to carry out evaluations, their motivation and commitment, the support provided by their unit/supervisor and their engagement in influencing other colleagues to achieve evaluative excellence.

Both internal evaluators and evaluation managers showed high self-confidence when talking about motivation and commitment for conducting evaluations. However, **evaluation managers perceived themselves less skilled in evaluation matters than internal evaluators, which might be aligned to the results on their lower participation on specific evaluation trainings** and that whilst there is a dedicated training for internal evaluators, there is no dedicated training for evaluation managers²⁸. When considering the support received on evaluation matters, evaluation managers felt more supported by their unit/supervisor than internal evaluators.

Figure 17. Proportion of stakeholders considering themselves somewhat or highly prepared to conduct evaluations based on the different items.



In any case, it should be noted that the interviews showed that evaluation procedures slightly change among regions as a result of the decentralized system of the Organization, and that the functions of evaluation managers, ROMEQ, and even the evaluators themselves, could change from one place to another.

²⁸ It must be noted that the Project Development and Project management trainings both have modules on evaluation, and the e-learning course on M&E has one module dedicated to evaluation managers.

9. EVALUATION PROMOTION AND USE

Key findings:

15. Despite staff involved in evaluation consider that evaluations somewhat reached all potential users, they believe the dissemination process should be further enhanced. For example, evaluation results are usually not made available to beneficiaries.
16. The dissemination of evaluation results has been enhanced in recent years by introducing tools such as the PRIMA for the Development Fund database, the evaluation repository and the evaluation briefs.
17. Internal staff considered that more clear guidelines and procedures for disseminating results should be made available and mainstreamed into the Organization's evaluation function.
18. Key factors facilitating evaluation use are the institutional efforts made in recent years toward promoting actionability of evaluation recommendations and the interventions OIG plans to carry out as part of its M&E strategy, such as the thematically-focused synthesis evaluations and the current results reporting assessment. The establishment of the new PRIMA for all system could play also a key role in enhancing evaluation dissemination and use.
19. Key factors hindering evaluation use are the need of additional management and institutional commitment in the promotion and use of evaluation results, the limited use of existing recommendations tracking mechanisms, the reduced engagement of stakeholders during the evaluation for ownership and the scarce resources available for evaluations.

Dissemination of evaluation results

As per IOM internal guidelines²⁹, all evaluation reports need to be shared with OIG, which will list the evaluations on the IOM website. In addition, OIG enters all evaluations into an internal repository of evaluation reports that are available for all staff since October 2019. Moreover, recent efforts toward promoting evaluation results include the introduction of evaluation briefs in mid-2018, as well as encouraging webinars, presentations and knowledge sharing activities, such as the participation in UN evaluation initiatives.

Internal evaluators and evaluation managers agreed on the fact that **results somewhat reached all potential users**. The evaluation managers shared the final evaluation reports in the vast majority of the cases with donors (84%) and regional offices (79%). More than half shared the reports with counterparts, governments and IOM Staff (OIG and other IOM offices). Beneficiaries were the one who receive the evaluation reports on a lesser extent, only in a 20% of the cases.

Despite having some general guidelines on how to proceed toward disseminating evaluation results in the IOM Project Handbook, all groups of informants expressed the need to **establish clear guidelines and common procedures to homogenize dissemination and use of results**. In that aspect, during the KII, programme managers were identified as responsible to

²⁹ IOM Project Handbook, Pag. 449, July 2017.

ensure the correct use of these guidelines as a way of promoting the work of the organization and implementing more efficient and effective interventions.

To the question if there is anything that should be done differently to enhance the use of evaluations, evaluation managers and internal evaluators identified that the **most important action to take is to enhance the results dissemination process** internally and externally. They pointed out that more efforts towards **sharing the results of previous evaluations**, as well as lessons learned and good practices, would have a positive impact in programming and implementation.

Use of evaluation results

As per IOM's Project Handbook, the use of evaluations is based on sustaining or replicating good practices and rectifying or avoiding lessons learned. Recommendations are expected to be based on these two central ideas, and the Organization has been putting considerable efforts in recent years toward ensuring an adequate follow up mechanism to measure the implementation of evaluation recommendations.

Factors enhancing evaluation use

- **Institutional efforts made toward promoting actionability of evaluation recommendations.** Evaluation managers and internal evaluators are of the opinion that the evaluation results somewhat led to decision making and that improved the quality of the interventions. In the same way, they mostly agreed on the assumption that both lessons learned and good practices were considered for other interventions. On the other hand, ROMEOS and other key stakeholder interviewed pointed out the need to keep strengthening the use of the management response tool in order to be more coherent and efficient in the use of results. The recently published management response and follow up guidance reflects these efforts on ensuring that evaluations are reviewed and considered for future planning and programming.
- **Feedback on uses given to evaluations,** since all surveyed groups agreed that they were informed about the use given to the evaluation. External evaluators openly manifested that they did not know if the evaluation recommendations were adequately followed up for their implementation, while the internal evaluators and evaluators managers agreed that they were followed up.
- **Strategies to be rolled out.** OIG has already identified to strengthen evaluation use, such as the planned synthesis evaluations on thematic area to inform decision-making, together with the meta-evaluation exercise. In addition, the establishment of the new PRIMA for all system could play also a key role in enhancing evaluation use and linking evaluation results to programming and planning.

Factors hindering evaluation use

- **The need for additional management and institutional commitment in the promotion and use of evaluation results.** Informants from all groups highlighted that more efforts for involving management and other key stakeholders are needed towards ensuring the

actionability of recommendations. In that line, the consistent proactive involvement of the Chiefs of Missions, regional directors and regional thematic experts was seen essential in reinforcing the use of evaluations. Moreover, evaluation managers have stressed the need to increase the internal communication to promote the use of the evaluation results.

- **Limited use of existing recommendations tracking mechanisms**, as pointed out by internal evaluators. In this line, internal and external evaluators stressed the importance of conducting strategic recommendations in order to be more efficient in their use and promotion of results. The promotion of the recently published management response guidelines could play a major role in promoting the use of management response matrix.
- **Little awareness on the utility of evaluations** within the project team. As a key informant expressed, “M&E has to be a culture, it needs to be an institutional position”. This was related to the idea that managers more sensitive about M&E will be able to negotiate with donors more financial resources for conduct these activities.
- **Reduced engagement of stakeholders to build ownership.** Internal and external evaluators highlighted the need to promote partner’s engagement and increase the participation of stakeholders during the evaluation process to accomplish better results and disseminate them among all partners involved. This is aligned with the results of the quality of evaluation reports, as limited participation of stakeholders was also pointed out in section on evaluation quality, component 5 on methodology, item 6. In this line, it was highlighted the need to focus on recommendations formulated for each implementing actor.
- **Scarce resources for evaluation.** The lack of human resources, together with the limited evaluation skills of staff was something mentioned frequently as a major impediment toward evaluation use. Evaluation managers have specified the importance for evaluators to have specific contextual and thematic knowledge of the object under evaluation. The three groups of informants agreed that the lack of financial resources and time available was a major impediment to accomplishing quality evaluations and disseminating results.

10. CONCLUSIONS

To what extent do IOM evaluation reports and their ToR meet the UNEG evaluation standards using evaluation criteria and including cross-cutting issues?

Conclusion 1. When it comes to meeting UNEG standards³⁰, the majority of **evaluation reports** satisfactorily presented what was evaluated and why (purpose and scope), how the evaluation was designed and conducted (evaluation questions, methodology and limitations), what was found and on what evidence (findings) and what was concluded from the findings in relation to main evaluation questions addressed. On the other hand, the majority of reports fell short in adequately addressing cross-cutting issues throughout the process, including comprehensive recommendations and including what, if anything, could be learned from the evaluations (lessons learned). Evaluation report templates have positively evolved over the years, as the self-evaluation report format included in the IOM Evaluation Guidelines 2006 is considerably simpler than the report template included in the IOM 2017 Project Handbook. In this line, some of the recommendations shared in the next section include a series of modifications for these templates to keep evolving.

Generally speaking, the great majority of **ToR** available meet the UNEG standards of including a clear description of the evaluation context and purpose, a description of the subject to be evaluated, and the scope of the evaluation and its objectives, including key evaluation criteria and questions. ToR also include a comprehensive description of the methodology to be considered for the evaluations as well as the key deliverables and workplan/timetable. On the other hand, roles and responsibilities are not consistently included in the majority of ToR, neither (and more importantly) are specific concerns on cross-cutting issues, particularly environmental issues and RBA.

The shortfalls found for evaluation reports and ToR might be explained by the fact that the evaluation reports and ToR templates included in the IOM Project Handbook do not address adequately these issues.

Conclusion 1 is linked to findings 5, 6, 7 and 13.

On which quality aspects or factors do the evaluation processes excel and where are they falling short?

Conclusion 2. During the **evaluation design**, the use of templates for drafting ToR and inception reports is widespread, which is very positive to ensure consistency. One of the key limitations found was the lack of a systematic approach to define whether an evaluation should be internal or external, as well as a standardized criterion to determine which evaluations are mandatory to be followed up by ROMEOS, who usually work on a demand-driven basis.

³⁰ As included in the document: United Nations Evaluation Group (2016). Norms and Standards for Evaluation. New York: UNEG 2017

As for the **implementation of the evaluation**, the overall key opportunities found are the strong institutional support toward the evaluation function and the richness of the emerging internal network of evaluators, which ensures that generated evaluation knowledge and skills built remain within the Organization. Moreover, it is important to recognize and support the role played by the internal evaluators who are performing their evaluation duties in addition to their other tasks.

Concerning major challenges, the limited funding available for evaluations, together with the short timeframes as perceived by evaluators and an incipient evaluation culture with limited understanding of the potential uses of evaluations within the Organization were considered key elements that affect negatively the quality of evaluations. Concerning specific processes, the selection of evaluation criteria and methodology was reported very effective and fit for purpose, as well as the communication flows with relevant stakeholders during the evaluation. On the other hand, key specific challenges during the implementation period include mainstreaming cross-cutting issues, the limited availability of M&E data, the diversity of internal evaluation skills, the difficulty in accessing key informants, and high staff turnover. In general, the feedback process is considered essential to ensuring quality and it is only negatively affected by the limited evaluation capacities of participants in some cases and the lack of a screening process for receiving comments.

Conclusion 2 is linked to key findings 8, 9 and 10.

Are evaluation reports and ToR developed based on the recommended templates provided by IOM guidelines?

Conclusion 3. The great majority of reports, especially internal evaluation reports, followed the structure of the template evaluation report included in the IOM Project Handbook of 2017. Similar is the case for the ToR. This had its advantages and disadvantages: following the templates ensured alignment to some of the key issues considered in the UNEG standards. Conversely, the reliance on these templates led to some of the shortfalls identified, with the consequence that those documents did not meet UNEG standards (see conclusion 1). Furthermore, it should be noted that **IOM guidelines include very useful templates beyond ToR and evaluation report templates that could increase the quality of the final evaluation products.** For example, IOM Project handbook include the evaluation matrix template, that is not consistently used and that has been proven to boost report quality. In addition, IOM guidelines include a comprehensive list of key questions suggested for incorporating cross-cutting themes into the evaluations, which has been hardly used in practice.

Conclusion 3 is linked to key findings 6 and 7.

Have the time and cost/resources assigned to the evaluations influenced the quality of the final product?

Conclusion 4. As included in the limitations section, it was not possible to analyse the influence of time and funds earmarked to evaluations and the way they influence evaluation quality. **Despite the fact that no imbalances in quality ratings were found between internal and**

external evaluations, and that higher evaluation budgets did not impact the quality of evaluation reports, the analysis is limited and no robust conclusions could be drawn.

For future reference, when considering the budget of internal evaluations, the time dedicated by internal evaluators should be factored in to establish the real cost associated to any given evaluation.

Conclusion 4 is linked to key finding 3.

What can be determined about the overall quality of the evaluation reports and where do the greatest opportunities for improvement lie?

Conclusion 5. The analysis on the quality of evaluation reports is based on the meta-evaluation tool specifically designed for the meta-evaluation. Some of the requirements included in the tool goes beyond the elements included in the UNEG checklist for evaluation reports and the templates included in the IOM Project Handbook. Therefore, the evaluations conducted in the period 2017-2019 did not have the reference quality checklist against which they have now been assessed. The consulting team present the results of the quality assessment being fully aware of the aforementioned situation.

One third of the evaluation reports assessed meet or exceed quality requirements considered. Roughly half of the evaluation reports meet certain requirements but not all of them, and around 20% of the reports did not meet the minimum requirements of quality established. Considering the relative weight given to the different components included in the meta-evaluation, **the greatest opportunities for improvement lie in the adequate inclusion of cross-cutting issues at all levels**, including specific evaluation questions and methodologies, addressing these issues in key findings, and more importantly, reflecting this analysis in the final set of recommendations. Similarly, **recommendations are also a key component on which to focus**, not only by ensuring cross-cutting issues are taken into consideration, but also clearly establishing their actionability (what, how and when) and responsibility (who is called upon to act). **The systematic inclusion of a specific section on lessons learned and good practices is also considered essential.**

Conclusion 5 is linked to key findings 1 and 5.

During the period covered, were there any changes in the quality of the evaluations conducted and if yes, what factors may have contributed towards these changes?

Conclusion 6. There is a positive evolution over time in the quality of evaluation reports. The analysis of the evolution of component rating over time shows that the rating for all of the components increased for internal evaluations during the period under assessment.

The reasons for this positive trend on the quality of reports are difficult to discern considering the limitations of the assessment, especially since the period of three years is limited to justify changes. On the one hand, the **introduction of ROMEOs** are clearly perceived as a major factor towards improving evaluation quality, as they are key experts providing guidance and support to the missions. However, these posts were introduced in 2016, previous to the period under assessment. In any case, it could be concluded that their involvement is becoming more fruitful

and their influence is reflected in the quality of reports over time. In addition, the **institutional efforts** made toward strengthening the evaluation function and the introduction of updated policies, manuals, guidelines and templates could also explain the consistent improvement of report quality. Finally, the introduction of the **internal evaluator training package** in late 2017 could have also considered to have a positive impact on the quality of evaluations, as further elaborated in conclusion 7.

Conclusion 6 is linked to key findings 2, 8, and 9.

What patterns of quality exist (if any) between those conducted by the trained internal evaluators, non-trained internal evaluators, and external evaluators? What strengths and gaps in evaluation competencies do the processes and reports indicate?

Conclusion 7. Generally speaking, there are no relevant differences on the quality of internal and external evaluation reports. Despite these similarities, evaluation reports conducted internally have experience a considerable positive evolution during the period 2017- 2019, something that was not identified for external evaluations. Additionally, **internal trained evaluators reach an overall rating considerably higher than the global average for both internal and external evaluators, which leads to conclude that efforts made in terms of capacity building appear to be paying off.** To complement this, surveyed staff expressed their satisfaction with the applicability of the training to their daily work, as well as their motivation and commitment. When it comes to gaps in evaluation competencies, evaluation managers have a lower perception that internal evaluators on their evaluation skills.

Conclusion 7 is linked to key findings 2, 3, 4, 12 and 14.

How were/are evaluations used and promoted? What factors facilitated or hindered the use of the evaluations?

Conclusion 8. When it comes to the **dissemination** of evaluation results, and as a general rule, there is no restriction on the public distribution of completed evaluation reports except those that relate to IOM transport services and resettlement programmes under contract with a specific government³¹. Upon donor agreement, evaluation reports are expected to be circulated not only to donors and host governments, but also among other key stakeholders, both internally and externally. Internally, the dissemination of evaluation results has been enhanced in recent years by introducing the evaluation briefs (2018), and the repository of evaluations for IOM staff (end of 2019) and will be further improved with the use of PRIMA for all database where evaluations will be stored. Concerning external stakeholders, all evaluations reports are publicly listed on the IOM evaluation website as per the evaluation policy and available upon request; however, this will soon be changed as all the evaluation reports will be accessible through a repository that will be available on a new webpage for IOM Evaluation. Despite this, all groups of stakeholders consulted agreed on the need to keep strengthening the evaluation dissemination process by providing more access to evaluation results, as well as promoting clearer guidelines and procedures for disseminating results. Only by ensuring that more efforts are done to increase the engagement of management, related staff (chiefs of mission, regional

³¹ IOM Project Handbook, page 449, July 2017.

directors and regional thematic experts) and other stakeholders, will have direct results in the accountability of the implementation of recommendations.

Concerning the **use of evaluations**, the key factors enhancing evaluation use are the institutional efforts made in recent years toward promoting the actionability of evaluation recommendations and the strategies OIG plans to carry out in the short term such as the thematically-focused synthesis evaluations and the current results reporting assessment. The key factors identified hindering evaluation use are the need for additional management and institutional commitment in the promotion and use of evaluation results, the limited use of tracking mechanisms for existing recommendations, the reduced engagement of stakeholders during the evaluation for ownership and the scarce resources available for evaluations. There is also a need to increase the internal communication and information knowledge management to improve the performance of the use and dissemination of evaluation's results.

Conclusion 8 is linked to key finding 15, 16, 17, 18 and 19.

11. GOOD PRACTICES

Widespread use of evaluation templates

The analysis reveals that a great majority of evaluation reports, and more specifically, internal reports, follow the templates included IOM Project Handbook and the IDF guidelines. This reveals a positive tendency on following the existing key documents as a main framework when conducting evaluations. Despite the fact that some of the templates are not consistently used, it is expected that changes in those that are followed would bring about positive changes in evaluation quality.

PRIMA IDF - an early and accessible evaluation data base

The Project Information and Management Application (PRIMA) is the system that facilitates the administration of the IOM Development Fund (IDF) and its projects. The PRIMA for IDF system was developed in as an initiative to deploy an institutional project information management solution to assist IOM to develop, manage, monitor, and report on IDF-funded projects and their results. More importantly for this assessment, all IDF related evaluations are uploaded on the PRIMA system, so it will be available to anyone with access to the system around the world. Despite the fact that the system is not a repository and does not provide direct access to evaluation reports, it made evaluations accessible, and eventually led to the development of the new PRIMA for all system. This system, which has only been rolled out since 2019, is expected to contribute to the dissemination of evaluations results by linking evaluation reports to project documents, which could be very useful when designing and planning interventions.

Evaluation coaching

The coaching system is part of the evaluation capacity building strategy launched by the Organization in recent years. Under this initiative, any internal evaluator can request the support and guidance of key experts and staff within the organization during the evaluation process. Although the first intention was to support internal evaluators on their first assignment, the initiative also gives the opportunity to any internal evaluator to request to support to colleagues with the goal of increasing evaluation quality. This good practice amplifies the communication channels among IOM evaluation staff around the world and helps disseminating knowledge and debates, while, consequently, fostering staff and organization cohesion. The coaching system has a great potential not only to strengthen the capacities of internal evaluators, but also to develop evaluation culture.

12. LESSONS LEARNED

A Meta-evaluation beyond the quality assessment

The scope of the current meta-evaluation went beyond assessing quality of evaluation reports (and ToR), and also put focus on evaluation processes and use. This has given a broader scope of analysis that helped understand and explain some of the key constraints and limitations identified as a result of the quality assessment conducted. The exercise did not only serve the purpose of controlling the quality of evaluations, but also the assessment of their adequacy and

opportunity to IOM context and their contribution to the achievement of the Organization's evaluation function. It provided a broader analysis and therefore it helped sustain and explain some of the findings with regard to the context and reality in which evaluations are conducted, ensuring a higher robustness for the whole analysis. All this should be considered for future meta-evaluation exercises, which should not only focus on analysing the evolution of evaluation report quality over time but also how processes and uses have changed compared to what has been presented in this meta-evaluation.

13. RECOMMENDATIONS

RECOMMENDATIONS TO ENHANCE EVALUATION REPORT QUALITY

Recommendation 1: IOM should **update the inception and evaluation report, as well as the ToR templates, and use them under all frameworks and donors.** The use of evaluation templates should remain mandatory for internal evaluators. As for external, it should be mandatory to include all specific sections considered in aforementioned templates. **The following items should be included in inception, evaluation and ToR templates:**

- **A standardized format to draft recommendations** should be added to the relevant section of the evaluation report template. The format presented in this recommendations section could be followed, as it includes the key elements considered toward ensuring the actionability and comprehensiveness of recommendations.
- **Include a specific section on lessons learned and good practices in evaluation report templates**, to ensure these are included in every evaluation report when applicable. Even when not applicable, the information on the lack of lessons learned and good practices would be also relevant to managers and other stakeholders. ILO templates for these items could serve as an inspiration³².
- Ensure that **all evaluations include an evaluation matrix** as a mandatory deliverable, and that reference is made to the relevant existing template in the IOM Project Handbook.
- Ensure **higher comprehensiveness** by including evaluation costs, duration of the evaluation, a list of tables and figures and an attached inception report in evaluation report templates. ToR templates should include a specific section on norms and standards, specifically request the identification of unintended effects and include management arrangements.
- Use the **meta-evaluation tool designed as a checklist** to ensure that all key aspects (items) are taken into consideration for all kind of evaluations. This will help increase the ratings dramatically in the next global meta-evaluation, as the tool should serve as a basis for the assessment of future evaluation reports.

Linked to: Conclusion 1, 3 and 5.

Resource implication: Not foreseen term

Responsibility: OIG with the support of ROMEOS if applicable.

Priority: Very high

Timeline: Short

³² https://www.ilo.org/global/docs/WCMS_165981/lang--en/index.htm

Recommendation 2: Evaluations would provide meaningful information about cross-cutting issues if gender-responsive actions, RBA and environmental sensitivity would be consistently included in all the phases of the evaluation. This should be considered from the early needs assessments, identification of stakeholders and definition of the evaluation methodology, to the final analysis and interpretation of results and when relevant, as part of the recommendations. In order to ensure that, it is proposed to:

- Include an **additional criterion on cross-cutting issues in the ToR template**, as well as in the template for evaluations questions. A link or reference to the template on cross-cutting thematic evaluation questions already included in the IOM Project Handbook should be also made for reference.
- Lesson learned and good practices templates should offer ideas and experiences on how to detect and reports cross-cutting themes.

Linked to: Conclusion 1 and 5

Priority: Very high

Resource implication: Not foreseen

Timeline: Short term

Responsibility: OIG, evaluation/programme managers, with the support of ROMEOs and thematic experts on RBA, gender and environment.

RECOMMENDATIONS TO STRENGTHEN THE EVALUATION PROCESS

Recommendation 3: In order to complement already mentioned institutional efforts made in recent years, the following strategies should be considered toward strengthening the evaluation process:

- Establish a **systematic approach to decide when to conduct internal or external evaluations**. The type and timing of an evaluation could depend on the amount of funding, thematic considerations, the purpose or the duration of the project under consideration, and should be negotiated with the donor. Criteria used by other UN agencies (i.e. UNDP, UNICEF) could be used as a reference.
- Keep promoting evaluative approaches and PPR for **evaluations with reduced budget, and in an overall sense, ensure that evaluations with low budgets focus on very specific areas or information needs**. Complex, multi-country or regional evaluations should ensure a minimum evaluation budget allowing sufficient time for its adequate implementation. Among the strategies to maximize the efficient use of evaluation resources would be to conduct joint thematic evaluations covering similar interventions under same or different donors and combine evaluation budgets.
- **Keep implementing evaluation capacity building activities and the coaching initiative**, enriching the existing network of internal evaluators and increasing evaluation skills of general staff as stated in the OIG M&E strategy 2018-2020. This will also contribute to develop a strong evaluation culture within the Organization and limit the negative impact of staff turnover.
- **Establish clear guidelines to homogenize the involvement of ROMEOs in different evaluations** (complexity, duration of the project, budget available, etc.), ensuring that their involvement is not only based on a demand-driven approach, but also according to some

shared criteria. This is particularly relevant when considering key evaluations for which ROMEOS in-depth involvement should be compulsory.

- **Keep developing and expanding a comprehensive and shared pool of evaluation experts**, both internal and external, including key thematic and geographical expertise, as well as working languages.
- **Recognizing the role played by internal evaluators.** The use of public recognition feeds, peer recognition and rewards could be complemented with other positive reinforcements, such as economic incentives, internal promotions, or allocating staff time to accomplish their evaluative function during key evaluation stages.

Linked to: Conclusion 2 and 6.

Priority: High

Resource implication: Not foreseen

Timeline: Mid-term

Responsibility: OIG, evaluation/programme managers, with the support of ROMEOS and thematic experts

RECOMMENDATIONS TO ENHANCE EVALUATION DISSEMINATION AND USE

Recommendation 4: IOM would enhance evaluation dissemination and use by considering the following:

- **Providing access to a** systematic classification of evaluation results in the evaluation repository, allowing the identification of key findings, conclusions, recommendations, lessons learned and good practices by thematic and geographical areas, type of evaluation, etc. would be key to boost internal evaluation use for decision making and programming.
- **Implementing strategies to ensure the inclusion and participation of the target population in the evaluation process** (e.g. include them in ToR, allocate mission time and budget for their participation, engage IOM staff in the inclusiveness of this target group). This could explore the relevance and strategic fit of the intervention outside the sphere of influence of donor/client/implementers, and make evaluations more according to RBA.
- Continue **raising awareness of the utility and uses of evaluation** among staff, as another strategy to build evaluation culture within the Organization. More specifically, efforts should be made to engage management, such as chiefs of mission, regional directors and regional thematic experts.

Linked to: Conclusion 8

Priority: High

Resource implication: Elaboration of the database. Awards granted.

Timeline: Mid-term

Responsibility: OIG, with the support of ROMEOS and thematic experts

14. LIST OF ANNEXES

Annex 1	Meta-Evaluation tool - Dimension 1 (Quality of evaluation reports)
Annex 2	Meta-Evaluation tool - Dimension 2 (Comprehensiveness of evaluation reports)
Annex 3	Meta-Evaluation tool - Dimension 3 (Comprehensiveness of ToR)
Annex 4	Questionnaire for Evaluation managers
Annex 5	Questionnaire for Internal Evaluators
Annex 6	Questionnaire for External Evaluators
Annex 7	Interview Protocol
Annex 8	List of documents reviewed
Annex 9	List of evaluation reports
Annex 10	Consistency issues considered
Annex 11	List of people interviewed
Annex 12	Detailed analysis of ratings per region
Annex 13	Inception report

ANNEX I –ANALYTICAL FRAMEWORK FOR 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	Project 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	Evaluator	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 project background	1. Purpose 2. Objectives	
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 BACKGROUND – Weight 5%		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
PB1. Brief outline of contextual factors	1. Socio-Economic context 2. Political context	
PB2. Clear and relevant description of key stakeholders	1. Rightholders 2. Implementing agency (ies) 3. Duty bearers/Responsibility holders	
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	
COMPONENT 3: EVALUATION BACKGROUND – Weight 5%		
ITEMS	ELEMENTS	RATING & COMMENTS
EB1. Explanation of purpose of evaluation	1. Purpose (s) (accountability, project improvement and/or organizational learning) 2. Explanation of evaluation purpose(s)	
EB2. Description of evaluation scope	1. Geographic coverage 2. Timeframe 3. Thematic coverage	
EB3. List of evaluation clients and main audiences of the report	1. Intended users (donors, implementing partners, etc.) 2. Intended use per client	
COMPONENT 4: EVALUATION CRITERIA & QUESTIONS – Weight 7,5%		
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	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
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 was justified.
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, Random Control Trial, Program Theory, Contribution analysis, Outcome mapping, participatory evaluation, etc.
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),
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. Environmental issues	The extent to which cross-cutting issues were considered 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.
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.
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.

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 purpose, questions and approach	
F2. Robustness	1. Findings justified by evidence 2. Disaggregation of data by key variables	Baselines and targets are used, or their omission is justified.
F3. Identification of causal factors leading to accomplishments and failures	1. The causal factors leading to achievement or non-achievement of results are clearly identified. 2. Description of unintended effects	
F4. IOM Cross-cutting issues are adequately addressed	1. Gender 2. Rights-based approach 3. Environmental issues	

COMPONENT 7: CONCLUSIONS – Weight 20%

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

COMPONENT 8: RECOMMENDATIONS – Weight 20%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
R1. Clarity	<ol style="list-style-type: none"> 1. Are clear and concise 2. Are based on findings and/or conclusions of the report 	<ol style="list-style-type: none"> 1. (i.e. one or two sentences followed by explanatory text, if needed) 2. Directly linked to findings and conclusions
R2. Relevance	<ol style="list-style-type: none"> 1. Address key issues and are useful, tied to the object and purposes of the evaluation. 2. Address cross-cutting issues 	
R3. Responsibility	<ol style="list-style-type: none"> 1. Specify who is called upon to act 2. Identify means for achievement 3. Specify priority or importance 	<ol style="list-style-type: none"> 1. Unit, department, office, etc. 2. Human, financial, material) 3. Low, medium, high
R4. Actionability	<ol style="list-style-type: none"> 1. Indicate the specific courses of action needed to remedy/or continue with the current situation 2. Recommend a time frame 	

COMPONENT 9: LESSONS LEARNED – Weight 2.5%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
LL1. Relevance	<ol style="list-style-type: none"> 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. Lessons learned are generalizations based on evaluation experiences that abstract from specific circumstances to broader situations.
LL2. Delimitation	<ol style="list-style-type: none"> 1. The lessons concisely capture the context from which they were derived. 2. LL target specific users 	
LL3. Applicability	<ol style="list-style-type: none"> 1. The lessons specifically suggest what should be repeated or avoided in future contexts to guide action. 2. The LL include causal factors. 	

COMPONENT 10: GOOD PRACTICES – Weight 2.5%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
GP1. Delimitation	<ol style="list-style-type: none"> 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	<ol style="list-style-type: none"> 1. The statements describe how the GP 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.	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.

ANNEX II – ANALYTICAL FRAMEWORK FOR 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	Project 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. TORs B. List of informants C. Inception report D. Data collection tools	Yes/No
C.17	Logically structured		Yes/No
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 FOR DIMENSION 3 - COMPREHENSIVENESS OF EVALUATION TERMS OF REFERENCE

DIMENSION 3 - COMPREHENSIVENESS OF TERMS OF REFERENCE			
Code	ITEMS	ELEMENTS	Rating
ToR.1	Background, context	Summarizes the overall context and the project or programme that will be evaluated.	Yes/No
ToR.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	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	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	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)	<ul style="list-style-type: none"> - Incorporate an assessment of relevant human rights and gender equality aspects through the selection of the evaluation criteria and questions; - Specify an evaluation approach and methods of data collection and analysis that are rights-based and gender-responsive; - Specify that evaluation data should be disaggregated 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 in order to answer the evaluation questions. The methodology should include approaches, data sources, techniques, participation levels, sampling and limitations	Yes/No
ToR.8	8. Deliverables	The ToR mention which deliverables are expected as products (reports, presentations, info graphs, 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

DIMENSION 3 - COMPREHENSIVENESS OF TERMS OF REFERENCE			
Code	ITEMS	ELEMENTS	Rating
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 consider 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
ToR.13	13. Estimated number of days and budget	Detail of number of days per evaluation team member and total budget, 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 – QUESTIONNAIRE FOR 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. Which is your current contract type at IOM?
 - "General"/G Staff
 - "National Officer"/NO Staff
 - "Professional"/P staff
 - "Director"/D staff
 - Service contract
 - Consultant
 - Intern
 - Other: _____

4. What is your experience in evaluation? (multiple choice)
 - As a programme manager
 - As a project developer
 - As a project assistant
 - As an M&E assistant
 - As a M&E officer
 - As a M&E focal point
 - Other: _____
 - No experience

5. In which region are you based?
 - South America (RO Buenos Aires)
 - European Economic Area and Switzerland (RO Brussels)
 - East Africa and the Horn of Africa (RO Nairobi)
 - North Africa and Middle East (RO Cairo)
 - North America, Central America and the Caribbean (RO San José)
 - West and Central Africa (RO Dakar)
 - Southern African (RO Pretoria)
 - Asia and the Pacific (RO Bangkok)
 - South-Eastern Europe (RO Vienna)
 - HQ (Geneva)

SECTION 1 – EVALUATION PROCESS AND INVOLVEMENT

1. How have you been involved in the evaluation of IOM projects and/or programmes?
(Mark all that apply).
 - I managed the evaluation (i.e. being the main focal point of the evaluation)
 - I managed both the programme being evaluated and its evaluation
 - I manage the program under evaluation, and someone else managed the evaluation
 - I was part of the programme team for the programme/project being evaluated
 - I was working in M&E
 - Others (please specify) _____

2. As an evaluation manager, with which type of evaluator have you worked with?? Mark all that apply
 - Internal evaluators
 - External evaluators
 - OIG evaluators

3. Who usually writes the TORs?
 - The evaluation manager/evaluation focal point
 - The programme manager (who is not the designated evaluation focal point)
 - The regional M&E officer
 - The central Evaluation office at OIG
 - Other (please specify) _____
 - I do not know

4. Who usually has final approval over the choice of evaluator(s?) - Multiple choice
 - IOM official responsible for the administrative and/or technical backstopping of the project
 - IOM evaluation manager
 - Regional office evaluation focal point
 - Central Evaluation office at OIG
 - Chief of Mission
 - Donor
 - Other, please specify _____
 - Do not know

5. After the evaluator/evaluation team is selected, are changes made to the Terms of Reference during the first phase of the evaluation (inception phase)? (Tick all that apply) – (Linked to question 2 of the Meta-evaluation).
 - No changes allowed
 - No changes are usually 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) _____

7. To whom is the draft evaluation report usually sent to for comments? [Mark all that apply].
 (Linked to question 2 of the Meta-evaluation))

- Project staff
 Relevant IOM staff in HQ
 Relevant IOM staff in country offices
 IOM Evaluation Office (OIG)
 Regional Monitoring and Evaluation Officers
 National counterparts
 Donor
 No one
 Other, please specify _____
 Do not know

8. In general, based on your experience, please rate the extent to which you consider comments received from stakeholders were valuable for improving the overall quality of the draft evaluation reports. (Linked to question 2 of the Meta-evaluation)

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

10. Could please elaborate on your response? (Linked to question 2 of the Meta-evaluation))

11. Which of the following cross-cutting issues are consistently considered in the evaluation process? Mark all that apply. (Linked to question 1 of the Meta-evaluation))

- Gender mainstreaming
 Rights based approach.
 Environmental sensitivity and sustainability.
 Other (specify).....

12. In case one or several of the aforementioned cross-cutting issues are not usually included, please explain the reasons. (Linked to question 1 of the Meta-evaluation)

13. In your experience, what are the most common challenges experienced during the evaluation process? (Please mark all that apply) Linked to question 2 of the Meta-evaluation)

- Lack of information about the intervention (M&E data, secondary data, etc.)
- Difficulties in data collection
- Language and cultural barriers
- Reliability of available data
- Communication with IOM staff
- Evaluation capacities and skills of IOM staff involved in process
- Evaluation capacities and skills of evaluator
- Communication with internal evaluators
- Communication with external evaluators
- The evaluation took too much time to complete
- There was not enough time to complete the evaluation
- Others (Please specify_____)

14. Please rate your overall level of agreement with the following statements (Linked to question 4 of the Meta-evaluation)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	N/A
Funding available for conducting evaluations is adequate	<input type="checkbox"/>						
Internal evaluators are easily identified for conducting the evaluation	<input type="checkbox"/>						
External evaluators are easily identified for conducting the evaluation	<input type="checkbox"/>						
Timeframe for completing the evaluation is usually adequate to produce quality products	<input type="checkbox"/>						
Evaluations are conducted in a timely manner , to be used for decision-making purposes	<input type="checkbox"/>						
Criteria and questions used for the evaluations are adequate	<input type="checkbox"/>						
Evaluation methodology is usually adequate for expressed purposes	<input type="checkbox"/>						

Communication with evaluator(s) during the evaluation is fluid	<input type="checkbox"/>						
---	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

15. What could be done to improve IOM’s evaluations?

SECTION 2 – EVALUATION CAPACITIES AND SKILLS

16. Have you received any specific evaluation training? (Linked to question 7 of the Meta-evaluation)

- 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: _____

17. If you received evaluation training provided by the IOM, how useful was it toward improving your work on evaluation? (Linked to question 7 of the Meta-evaluation)

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

18. During the evaluation process, how frequently do you use/consult the following evaluation documents? (Linked to question 7 of the Meta-evaluation)

	Never	Rarely	Sometimes	Often	Always	Do not know the resource
IOM Project Handbook	<input type="checkbox"/>					
IOM Evaluation Guidelines	<input type="checkbox"/>					
IOM M&E Brief Guidance	<input type="checkbox"/>					
Guidance for Addressing Gender in Evaluations	<input type="checkbox"/>					
IDF Evaluation guidelines	<input type="checkbox"/>					
Internal Evaluator training materials	<input type="checkbox"/>					

	Never	Rarely	Sometimes	Often	Always	Do not know the resource
M&E e-learning Course materials	<input type="checkbox"/>					
Others (Please specify _____)	<input type="checkbox"/>					

19. To what extent do you consider yourself prepared to manage evaluations as part of your tasks? (Linked to question 7 of the Meta-evaluation)

Elements	Level of preparedness			
	None	Little	Some	High
Skilled	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Motivated (requires willingness)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Committed (involves continuous action).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supported by unit/supervisors	<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 PROMOTION AND USE

20. With whom do you usually share the final evaluation report? (multiple choice)

- Counterparts
- Donors
- Government
- Beneficiaries
- Other IOM offices
- OIG
- Regional offices
- No one
- Other (please specify) _____

21. Please rate your level of agreement with the following statements (Linked to question 8 of the Meta-evaluation)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	Don't know
I am informed of the use given to the evaluation results	<input type="checkbox"/>						
Evaluation(s) recommendations are adequately followed up for their implementation	<input type="checkbox"/>						

Evaluation reports are adequately promoted	<input type="checkbox"/>						
The evaluation(s) results reach all potential users	<input type="checkbox"/>						
The evaluation(s) lead to decision making	<input type="checkbox"/>						
The evaluation(s) improve the quality of the interventions	<input type="checkbox"/>						
Identified good practices are considered for other interventions	<input type="checkbox"/>						
Identified lessons learned are considered for other interventions	<input type="checkbox"/>						

22. Is there anything that should be done differently to enhance the use of evaluations?

ANNEX V – QUESTIONNAIRE FOR INTERNAL EVALUATORS

DEMOGRAPHICS

1. Can you please enter your 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. Which is your current contract type at IOM?
 - "General"/G Staff
 - "National Officer"/NO Staff
 - "Professional"/P staff
 - "Director"/D staff
 - Service contract
 - Consultant
 - Intern
 - Other

4. What is your experience in evaluation? (Multiple choice)
 - As a programme manager
 - As a project developer
 - As a project assistant
 - As an M&E assistant
 - As a M&E officer
 - As a M&E focal point
 - No experience
 - Other: _____

5. Years conducting evaluations
 - Less than 1 year
 - 1 to 3 years
 - 4 to 6 years
 - 7 to 9 years
 - 10 and more years

6. How many evaluations have you conducted for the IOM?
 - One evaluation
 - 2 - 3 evaluations
 - More than 3 evaluations

7. What was your role during the evaluation conducted for the IOM? (multiple choice)

- Sole evaluator
- Team leader
- Team member
- Other (please specify) _____

8. Evaluation related studies

- No formal training on evaluation
- Evaluation trainings (less than 50 hours)
- Evaluation training (more than 50 hours)
- Master in evaluations
- Evaluation related PhD

SECTION 1 – EVALUATION PROCESS AND INVOLVEMENT

1. How did you get involved in the evaluation process?

- Open call through the internal evaluator roster
- Direct assignment /allocation
- Others (please specify) _____

2. Concerning the terms of reference, after the evaluator is selected, what changes are usually made during the evaluation inception phase? – (Linked to question 2 of the Meta-evaluation)

- 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. As far as you know, to whom are the deliverables (inception report, draft evaluation reports, etc.) usually sent to for comments? [Please mark all that apply]. (Linked to question 2 of the Meta-evaluation)

- Project staff
- Relevant IOM staff in HQ
- Relevant IOM staff (specialists etc.) in regional/country offices
- IOM Evaluation Office (OIG)
- Regional monitoring and evaluation officers
- National counterparts
- Donor
- No one
- Other, please specify _____

Do not know

4. Please rate the extent to which you consider comments received were valuable for improving the overall quality of draft evaluation reports. (Linked to question 2 of the Meta-evaluation)

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

5. Could please elaborate on your response? (Linked to question 2 of the Meta-evaluation)

6. Which of the following cross-cutting issues are consistently considered in the evaluation process? Tick all that apply. (Linked to question 1 of the Meta-evaluation)

- Gender mainstreaming
 Rights based approach
 Environmental sensitivity and sustainability
 Other (specify).....

7. In case one or several of the aforementioned cross-cutting issues are not usually included, please explain the reasons. (Linked to question 1 of the Meta-evaluation)

8. In your experience, what are the common challenges experienced during the evaluation process? (Multiple choice). Linked to question 2 of the Meta-evaluation)

- Lack of information about the intervention (M&E data, secondary data, etc.)
 Difficulties accessing key informants
 Language and cultural barriers
 Reliability of available data
 Communication with IOM staff
 Evaluation capacities and skills of IOM staff
 Others (Please specify_____)

9. Please rate your level of agreement with the following statements (Linked to question 4 of the Meta-evaluation)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree
Funding available for conducting evaluations is adequate	<input type="checkbox"/>					
Timeframe for completing the evaluation is adequate to produce quality products	<input type="checkbox"/>					
Evaluations are conducted in timely fashion to be used for decision-making purposes	<input type="checkbox"/>					
Criteria and questions used for evaluations are adequate	<input type="checkbox"/>					
Evaluations methodology is usually adequate for expressed purposes	<input type="checkbox"/>					
Internal communication during the evaluations is fluid	<input type="checkbox"/>					

10. What could be done to improve IOM’s evaluations?

SECTION 2 – EVALUATION CAPACITIES AND SKILLS

11. Have you received any specific evaluation training within IOM? (Linked to question 7 of the Meta-evaluation)

- Yes
- No

If you received specific evaluation training within IOM, please select which training(s) (Linked to question 7 of the Meta-evaluation)

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

12. If you received evaluation training provided by the IOM, how useful was it toward improving your work on evaluation? (Linked to question 7 of the Meta-evaluation)

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

13. During the evaluation process, how frequently do you use/consult the following evaluation documents? (Linked to question 7 of the Meta-evaluation)

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

14. To what extent do you consider yourself prepared to conduct evaluations as part of your tasks? (Linked to question 7 of the Meta-evaluation)

Elements	Level of preparedness			
	None	Little	Somewhat	High
Skilled	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Motivated (requires willingness)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Committed (involves continuous action).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supported by unit/supervisors	<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 PROMOTION AND USE

15. Please rate your level of agreement with the following statements (Linked to question 8 of the Meta-evaluation)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	Don't know
I was informed of the use given to evaluations	<input type="checkbox"/>						
Evaluations recommendations were	<input type="checkbox"/>						

adequately followed up for their implementation							
Evaluation reports were adequately promoted	<input type="checkbox"/>						
The evaluation(s) results reached all potential users	<input type="checkbox"/>						
The evaluation(s) led to decision making	<input type="checkbox"/>						
The evaluation(s) improved the quality of the interventions	<input type="checkbox"/>						
Identified good practices were considered for other interventions	<input type="checkbox"/>						
Identified lessons learned were considered for other interventions							

16. Is there anything that should be done differently to enhance the use of evaluations?

ANNEX VI – QUESTIONNAIRE FOR EXTERNAL EVALUATORS

DEMOGRAPHICS

1. Can you please enter your gender?
 - Female
 - Male
 - Other gender identities

2. Evaluation related studies
 - No formal training on evaluation
 - Evaluation trainings (less than 50 hours)
 - Evaluation training (more than 50 hours)
 - Master in evaluations
 - Evaluation related PhD

3. Years conducting evaluations
 - 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 the IOM?
 - One evaluation
 - 2 - 3 evaluations
 - More than 3 evaluations

5. What was your role during the evaluation conducted for the IOM? (multiple choice)
 - Sole evaluator
 - Team leader
 - Team member
 - Other (please specify) _____

SECTION 1 – EVALUATION PROCESS AND INVOLVEMENT

1. How did you get involved in the evaluation process?
 - Open call through the internal evaluator roster
 - Direct assignment /allocation
 - Directly approach to conduct the evaluation
 - Others (please specify) _____

2. Concerning the terms of reference, after the evaluator is selected, what changes are usually made during the evaluation inception phase? – (Linked to question 2 of the Meta-evaluation)

- 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. As far as you know, to whom are the deliverables (inception report, draft evaluation reports, etc.) usually sent to for comments? [Please mark all that apply]. (Linked to question 2 of the Meta-evaluation)

- Project staff
- Relevant IOM staff in HQ
- Relevant IOM staff (specialists etc.) in regional/country offices
- IOM Evaluation Office (OIG)
- Regional monitoring and evaluation officers
- National counterparts
- Donor
- No one
- Other, please specify _____
- Do not know

4. Please rate the extent to which you consider comments received were valuable for improving the overall quality of draft evaluation reports. (Linked to question 2 of the Meta-evaluation)

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

5. Could please elaborate on your response? (Linked to question 2 of the Meta-evaluation)

6. Which of the following cross-cutting issues are consistently considered in the evaluation process? Tick all that apply. (Linked to question 1 of the Meta-evaluation)

- Gender mainstreaming
- Rights based approach

- Environmental sensitivity and sustainability
- Other (specify).....

7. In case one or several of the aforementioned cross-cutting issues are not usually included, please explain the reasons. (Linked to question 1 of the Meta-evaluation)

8. Which of the following document were shared with you to support the evaluation process? (Multiple choice) (Linked to question 7 of the Meta-evaluation)

- IOM Project Handbook
- IOM Evaluation Guidelines
- IOM M&E Brief Guidance
- Guidance for Addressing Gender in Evaluations
- IDF Evaluation guidelines (if applicable)
- Others (Please specify_____)
- None

9. In your experience, what are the common challenges experienced during the evaluation process? (Multiple choice). (Linked to question 2 of the Meta-evaluation)

- Lack of information about the intervention (M&E data, secondary data, etc.)
- Difficulties accessing key informants
- Language and cultural barriers
- Reliability of available data
- Communication with IOM staff
- Evaluation capacities and skills of IOM staff
- Others (Please specify_____)

10. Please rate your level of agreement with the following statements (Linked to question 4 of the Meta-evaluation)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree
Funding available for conducting evaluations was adequate	<input type="checkbox"/>					
Timeframe for completing the evaluation was adequate to produce quality products	<input type="checkbox"/>					
Evaluation was conducted in timely fashion to be used for decision-making purposes	<input type="checkbox"/>					

Criteria and questions used for evaluations were adequate	<input type="checkbox"/>					
Evaluation methodology was adequate for expressed purposes	<input type="checkbox"/>					
Communication with IOM during the evaluations was fluid	<input type="checkbox"/>					

11. What could be done to improve IOM’s evaluations?

SECTION 3 – EVALUATION PROMOTION AND USE

12. Please rate your level of agreement with the following statements (Linked to question 8 of the Meta-evaluation)

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly Agree	Don’t know
I was informed of the use given to evaluations	<input type="checkbox"/>						
Evaluations recommendations were adequately followed up for their implementation	<input type="checkbox"/>						
Evaluation reports were adequately promoted	<input type="checkbox"/>						
The evaluation(s) results reached all potential users	<input type="checkbox"/>						
The evaluation(s) led to decision making	<input type="checkbox"/>						
The evaluation(s) improved the quality of the interventions	<input type="checkbox"/>						
Identified good practices were considered for other interventions	<input type="checkbox"/>						
Identified lessons learned were considered for other interventions							

13. Is there anything that should be done differently to enhance the use of evaluations?

ANNEX VII – INTERVIEW PROTOCOL

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

QUESTION: What are the key processes and logic followed when preparing evaluations? What is the involvement of OIG/Evaluation, regional office Evaluation or country representatives in the evaluation design? How could that involvement be strengthened/improved?

AREA: Decision making with regards to ToR/ Evaluation design/ External teams

GUIDING QUESTIONS:

- What is the basis (why and how) to conduct an evaluation?
- How are evaluation timeframes and timing decided?
- What is the basis to conduct internal or external evaluations?
- Who elaborates the ToR?
- What is your level of involvement/participation in all this process?

2. SECTION: During the evaluation (considers questions 2, 3, 4 and 7 of the meta-evaluation))

QUESTION: To what extent are you involved in evaluations?

AREA: Participation/ Process/ Supervision

GUIDING QUESTIONS:

- Were you consulted during the evaluation process?
- Who supervises the evaluation process?
- 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 do you think could be done to improve the evaluation process?
- What are the main limitations to conduct evaluations (resources, skills and competencies, time allocated, etc.)?
- To what extent on-going efforts in M&E had an impact on evaluation products?

3. SECTION: After the evaluation: Use and utility of evaluations (considers questions 5, 6 and 8 of the meta-evaluation)) (considers questions 5, 6 and 8 of the meta-evaluation))

QUESTIONS: To which extent are evaluation reports useful? AREA: Uses/ Dissemination

GUIDING QUESTIONS:

- Who receives the evaluation reports?
- What use is given to evaluations? Are they useful toward decision making?
- Are evaluation reports realistic? Are recommendations realistic?
- How evaluation quality is evolving over time? What are the key contributing factors?
- Is there any kind of rethinking of interventions?

- Could you mention any lessons learned from the evaluations? Were they incorporated in further/other interventions?
- Are there any dissemination procedures and recommendation follow up mechanisms in place?
- What could be done to improve the use/uptake of evaluations? What is the perceived credibility of evaluations?

ANNEX VIII – 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
-	IOM Development Fund Project Evaluations
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
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
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
2018	UNICEF GEROS Meta-Analysis 2017
2018	UN Women Meta evaluation 2017 (GERAAS)

ANNEX IX– LIST OF EVALUATIONS

The following is the summary list of the evaluation reports assessed. For more detailed information, and excel sheet with all key relevant data is attached.

ID	Year	Title	Region(s)	Evaluator	Overall Rating	OVERALL RATING
1	2018	A Review of Knowledge Management in IOM: Current Status and Future Perspectives	Global	Internal	2,44	Approach requirements
2	2018	Project Evaluation Report - Building Capacity of Cambodia to Manage Outward Migration (BECOME)	Asia and the Pacific	Internal	3,19	Meets requirements
3	2018	IOM's Private Sector Partnership Strategy 2016–2020: A Thematic Evaluation	Global	Internal	3,06	Meets requirements
4	2018	Final Evaluation of IOM's Disaster Preparedness for Effective Response Project (PREPARE)	Asia and the Pacific	Internal	3,24	Meets requirements
5	2017	Improving Access to Health, Water and Sanitation Services, Education, And Freedom from Violence in Cox's Bazar	Asia and the Pacific	External	2,47	Approach requirements
6	2019	Internal program evaluation of community-based approaches to PVE in Bosnia and Herzegovina including two projects: "Community Based Approaches to Support Youth in Targeted Municipalities in BiH" and "Moving Towards Sustainable Approaches to Prevent Violence"	Europe	Internal	4,01	Exceeds requirements
7	2018	Independent Final Evaluation of the Project "Support to Strengthening Migration and Asylum Management in Serbia"	Europe	External	2,69	Approach requirements
8	2017	Final internal evaluation of the project "Support to Border Police in Addressing Increased Migration Flows along the Western Balkan Route"	Europe	Internal	3,39	Meets requirements
9	2018	Assessing the climate change, environmental degradation and migration nexus in South Asia.	Asia and the Pacific	Internal	3,25	Meets requirements
10	2018	Ex-post internal evaluation of the project "Increasing Capacities for Addressing Irregular Migration in Serbia"	Europe	Internal	4,02	Exceeds requirements
11	2019	Upholding the Rights of Migrant Workers in Special Economic Zones (SEZs) by Ensuring Supply Chains are Exploitation Free: Viet Nam, Cambodia and Lao People's Democratic Republic	Asia and the Pacific	External	2,65	Approach requirements
12	2018	Technical Support to Global Consultation on Migrant Health – Sri Lanka	Asia and the Pacific	External	2,88	Approach requirements
13	2019	Promoting Development-Friendly Migration Policies in Viet Nam through Enhancing the Evidence Base for Policy	Asia and the Pacific	Internal	2,94	Approach requirements

ID	Year	Title	Region(s)	Evaluator	Overall Rating	OVERALL RATING
14	2019	A Thematic Evaluation of The IOM's Migration Crisis Operational Framework (MCOF)	Global	Internal	2,97	Approach requirements
15	2018	Real Time Evaluation of IOM's Response to the Rohingya Crisis	Asia and the Pacific	External	3,21	Meets requirements
16	2018	Project Evaluation Report M.0290: Enhanced Skills Development and Qualification Recognition of Labour Migrants from Bangladesh	Asia and the Pacific	Internal	3,41	Meets requirements
17	2019	Community Based Tribal Conflict Mitigation and Peace Building in Enga and Morobe Province, Papua New Guinea (COMBAT-COMIT) Phase I and II	Asia and the Pacific	Internal	3,45	Meets requirements
18	2018	Mid-Term Evaluation 2016-2018 "Tajik-Afghan Integration, Resilience And Reform Building Programme" - TAIRR	Asia and the Pacific	External	1,95	Missed requirements
19	2018	Consolidation of Sri Lanka Biometrics Project: Biometrics Phase 3	Asia and the Pacific	Internal	2,72	Approach requirements
20	2018	Development of a National Framework for Durable Solutions in Vanuatu	Asia and the Pacific	Internal	2,95	Approach requirements
21	2018	IOM'S L3 Procedures and Practices Through the Lens of the Rohingya Response	Asia and the Pacific	External	1,59	Missed requirements
22	2019	Ex-post internal evaluation of the project "Piloting Temporary Labour Migration of Georgian Workers to Poland and Estonia"	Europe	Internal	3,59	Meets requirements
23	2017	Review of Migration Policy Funded Projects 2007-2016	Global	Internal	2,00	Missed requirements
24	2017	External Evaluation of an IDF Regional Project: Strengthening Capacity for the Promotion and Protection of Human Rights of Migrants in a Vulnerable Situation in South America (PRODEHSA)	America & Caribbean	External	3,07	Meets requirements
25	2018	Proyecto IDF "Fortaleciendo las capacidades de los países del Mercado Común del Sur (MERCOSUR) sobre: trata de personas y mujer migrante" CT. 0975	America & Caribbean	Internal	3,23	Approach requirements
26	2017	Evaluation of The Ghana Integrated Migration Management Approach Project	Africa	Internal	2,59	Approach requirements
27	2018	Rapport d'Evaluation Renforcement Des Capacites Pour Une Gestion Strategique Des Ressources De La Diaspora De La Rd Congo	Africa	Internal	2,22	Approach requirements
28	2018	Post Project evaluation of the project "Improving Capacities to Leverage Remittances for development In Ghana"	Africa	Internal	2,01	Approach requirements
29	2017	Evaluation finale du projet Enhancing Border Management to Counter Terrorism in Burkina Faso	Africa	Internal	1,42	Missed requirements
30	2017	Mid-term evaluation of the project Combating trafficking in human beings in Burkina Faso through the strengthening of the national framework and capacities in identifying victims and improving data collection	Africa	Internal	1,96	Missed requirements

ID	Year	Title	Region(s)	Evaluator	Overall Rating	OVERALL RATING
31	2017	Rapport d'évaluation du projet SIRIRI	Africa	Internal	3,09	Meets requirements
32	2017	Evaluation: Camp Coordination & Camp Management (CCCM)	Africa	External	2,28	Approach requirements
33	2017	Evaluation du projet « Améliorer la capacité collective de gestion des frontières et protéger les communautés de frontière entre la Mauritanie et le Mali »	Africa	External	2,27	Approach requirements
34	2018	Final External Evaluation of the project "Ghana Integrated Migration Management Approach" (GIMMA)	Africa	External	2,42	Approach requirements
35	2019	Impact migrants as messengers. The impact of peer to peer communication on potential migrants in Senegal	Africa	Internal	2,08	Approach requirements
36	2018	Evaluation of the project "contributing to the achievement of Sustainable Development Goals related to migration management in the Dominican Republic	America & Caribbean	Internal	2,81	Approach requirements
37	2019	Ex-post Project Evaluation Report "Enhancing Migration Management in Belize through an upgrade of the Migration Information Data Analysis System (MIDAS)"	America & Caribbean	Internal	2,75	Approach requirements
38	2017	Informe de evaluación final de Proyecto "Prevención y reintegración de jóvenes en riesgo en la costa Caribe de Nicaragua"	America & Caribbean	External	2,66	Approach requirements
39	2018	Evaluación ex-post del proyecto iniciativa de gestión de información de movilidad humana en el triángulo norte - NTMI	America & Caribbean	External	2,41	Approach requirements
40	2018	Rapport d'évaluation finale Projet "Aide aux enfants et femmes vulnérables dans les zones frontalières entre Haïti et la République Dominicaine"	America & Caribbean	External	3,12	Meets requirements
41	2018	Final Evaluation Report, Strengthening the Institutional Capacity of Criminal Justice Actors in the Caribbean to Counter Trafficking in Persons Using a Victim Centered Approach Project	America & Caribbean	External	2,77	Approach requirements
42	2017	IOM-led project Migration, Environment and Climate Change: Evidence for Policy (MECLEP)	Global	External	2,93	Approach requirements
43	2017	Final Report - Evaluating the Effectiveness of Regional Migration Program Models on Providing Assistance to Vulnerable Migrants	Global	External	2,33	Approach requirements
44	2019	Evaluación interna del proyecto IDF "fortalecimiento del enfoque de derechos humanos en las políticas migratorias de los estados parte del MERCOSUR y estados asociados	America & Caribbean	Internal	3,29	Meets requirements
45	2017	Report on IOM Development Fund 2017 Ex-Post Evaluations	Global	Internal	0,88	Missed requirements
46	2019	Fourth Evaluation of the IOM Development Fund	Global	Internal	2,08	Approach requirements
47	2018	Formation professionnelle et opportunité des moyens de subsistance pour les migrants régularisés au Maroc	Africa	Internal	1,92	Missed requirements

ID	Year	Title	Region(s)	Evaluator	Overall Rating	OVERALL RATING
48	2017	Impact Evaluation of IOM's Assisted Voluntary Return and Reintegration (AVRR) Programme - FINAL REPORT	Africa	External	1,86	Missed requirements
49	2018	On shelter interventions in Anbar Governorate, Iraq OFDA Project	Africa	Internal	1,63	Missed requirements
50	2018	Final Evaluation of the Project "Addressing the needs of Stranded and Vulnerable Migrants in Targeted Sending, Transit and Receiving Countries"	Africa	External	2,67	Approach requirements
51	2019	Education, Training and Knowledge-Sharing on Migration in Tunisia	Africa	Internal	3,23	Meets requirements
52	2019	Reinforce Capacities to Implement the Counter-Trafficking Legislation in Tunisia - Share II phase two	Africa	External	3,07	Meets requirements
53	2017	Evaluation of the Voluntary Return Assistance in Libya	Africa	External	2,10	Approach requirements
54	2017	Evaluation Report of The Project Strengthening Security In Djibouti And Safety Of Migrants Through Counter-Trafficking, Marine Safety And Emergency Assistance	Africa	External	1,72	Missed requirements
55	2017	Evaluation of International Organization for Migration (IOM) South Sudan Health and WASH response: 2014-2016	Africa	External	3,16	Meets requirements
56	2018	Strengthening Labour Migration Management in Madagascar (Project Number LM. 0283) - Report on an Internal Ex-Post Project Evaluation	Africa	Internal	2,00	Missed requirements
57	2017	Malawi Migration Profile Project - Project Evaluation Report 2017	Africa	Internal	1,95	Missed requirements
58	2017	Namibia Migration Profile Project, Project Evaluation Report	Africa	Internal	1,89	Missed requirements
59	2017	Addressing Irregular Migration in Southern Africa - An Internal Review of Project Achievements (2010-2016)	Africa	Internal	2,81	Approach requirements
60	2017	Final Report Voices from the Underground: End of Project Evaluation 2014-2016	Africa	External	2,63	Approach requirements
61	2017	Evaluation of the TB in Mining Sector Project	Africa	External	2,97	Approach requirements
62	2017	Camp Coordination and Camp Management and Disaster Risk Management Capacity Building Programme in Southern Africa: Phase III	Africa	External	2,16	Approach requirements
63	2018	Strengthening coordination to respond to trafficking in persons and ensure justice and protection for victims of trafficking in Mozambique project	Africa	Internal	2,36	Approach requirements
64	2018	Final Project Evaluation "Building Sustainable Peace and Promoting Human Security of Cross-border Communities and Mobile Populations through Integrated Border Security and Management and Community-Based Peacebuilding Activities"	Africa	External	2,94	Approach requirements

ID	Year	Title	Region(s)	Evaluator	Overall Rating	OVERALL RATING
65	2018	Final Evaluation Report "Capacity-Building in Integrated Border Management in Southern Mozambique" (TC.0861)	Africa	External	2,65	Approach requirements
66	2019	Capacity-Building for Diaspora Engagement in Mozambique (CE.0298) Ex-Post Evaluation Report, 2018	Africa	Internal	1,44	Missed requirements
67	2019	Improving Labour migration management in Belize	America & Caribbean	Internal	2,55	Approach requirements
68	2017	Evaluation of IOM's Approach for Providing Assistance and Protection to Migrant Children, including Unaccompanied and Separated Ones.	Global	Internal	3,26	Meets requirements
69	2017	Midterm Evaluation of IOM Gender Equality Policy 2015-2019	Global	Internal	2,05	Approach requirements
70	2018	A Thematic Evaluation Of The Displacement Tracking Matrix (DTM)	Global	Internal	2,46	Approach requirements
71	2018	Final Report - Mid Term Evaluation MIDA FINNSOM Health and Education Project Somalia	Africa	Internal	2,77	Approach requirements
72	2018	Evaluation de mi-parcours du projet "Migrant Response and Resource Mechanism (MRRM) Phase II"	Africa	Internal	1,49	Missed requirements
73	2019	Mid-term Internal Evaluation of the project: "Appui au renforcement de la gestion de la Migration au Niger" (Strengthening Migration Management in Niger)	Africa	Internal	2,61	Approach requirements
74	2019	Mecanisme De Reponse Et Ressources Pour Les Migrants (Phase Ii)	Africa	Internal	2,95	Approach requirements
75	2019	Strengthening Labour Migration Management in Lesotho	Africa	Internal	1,78	Missed requirements
76	2019	Strengthening Protection for Vulnerable Populations and Improving access to Psychosocial Support to Displaced Children in Emergency Settings in Mozambique	Africa	Internal	1,88	Missed requirements
77	2019	SRHR-HIV Knows No Borders Project	Africa	External	2,50	Approach requirements
78	2017	Final Internal Evaluation of the project "Strengthening Surveillance and Bilateral Coordination Capacity along the Common Border between Belarus and Ukraine (SURCAP Phase II)"	Europe	Internal	2,81	Approach requirements
79	2018	Towards an Effective Migrants' Right to Health in Central Asia: Assisting Governments in Enhancing the Provision of Health Services for Migrants - Kyrgyzstan	Asia and the Pacific	Internal	3,80	Meets requirements
80	2017	Evaluation of the EU-Community Stabilisation Programme In Kosovo	Europe	External	3,48	Meets requirements
81	2018	Evaluation Report for final internal evaluation of project "Combating Trafficking in Persons in Georgia and Addressing the Demand Side of Trafficking"	Europe	Internal	3,07	Meets requirements

ID	Year	Title	Region(s)	Evaluator	Overall Rating	OVERALL RATING
82	2018	Mid-term internal evaluation of the project "Technical assistance to the Government of Kyrgyzstan to strengthen the legal and operational framework for combating trafficking in persons"	Europe	Internal	3,49	Meets requirements
83	2018	Final internal independent evaluation of the project "Integration and Stabilization Support through Livelihoods for IDPs and the Conflict-Affected Population in Ukraine"	Europe	Internal	3,51	Meets requirements
84	2018	Final Internal Evaluation of the project "Addressing Migration Movements through Capacity Building in Central Asia - Phase III" - Internal Only Version	Asia and the Pacific	Internal	3,08	Meets requirements
85	2018	Final Evaluation Report Project "Preventing unsafe migration from Albania towards European Union member states" Campaign "Choose opportunities: not irregular migration!"	Europe	External	2,00	Missed requirements
86	2019	Final External Evaluation of the "Enhancement of National Capacities to Combat Human Trafficking in Azerbaijan (ENCT)" Project and "Supporting the Enhancement of National Capacities to Combat Human Trafficking in Azerbaijan (SNCT) Project	Europe	External	2,47	Approach requirements
87	2019	Expost Evaluation of the project "Strengthening Capacities of Armenian National Security Service and Border Guards in their Response to Migration Crisis"	Europe	Internal	3,22	Meets requirements
88	2019	Final External Evaluation of the "Enhancing Cooperation Measures to effectively Combat Trafficking in Persons through Capacity Building and Technical Assistance in Azerbaijan – phase VI (ECMCT)" Project	Europe	External	2,76	Approach requirements

ANNEX X – STRATEGIES TO ENSURE CONSISTENCY

GENERAL ISSUES

1. **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) of 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.
2. **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.
3. 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).
4. In the scoring tool, “no data” was entered when information was not available and N/A when not applicable.
5. **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).

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).

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.

COMPONENT 2: PROJECT BACKGROUND – Weight 5%

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.

COMPONENT 3: EVALUATION BACKGROUND – Weight 5%

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. Timeframe 3. Thematic coverage	Consider implicit information: if it is clear that it is a final evaluation but no timeframe 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.

COMPONENT 4: EVALUATION CRITERIA & QUESTIONS – Weight 7,5%

ITEMS	ELEMENTS	COMMENTS
-------	----------	----------

CQ1. Declaration of the evaluation criteria	<ol style="list-style-type: none"> 1. Evaluation criteria 2. Justification for their use 	<p>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.</p> <p>OECD 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.</p> <p>If no mention to evaluation criteria, but evaluation criteria is implicitly included within evaluation questions, considered the criteria included, but not complete.</p>
CQ2. Relevance of evaluation questions	<ol style="list-style-type: none"> 1. Evaluation questions addressed the goals and purpose of the exercise. 2. Questions include cross-cutting issues 	<p>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.</p> <p>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).</p>
CQ3. Inclusion of an evaluation matrix	<ol style="list-style-type: none"> 1. Evaluation matrix, including evaluation indicators 2. Methods and sources of data per question included in evaluation matrix 	<p>Reference indicators and benchmarks should be included, as this is how the judgment will be justified.</p> <p>If Evaluation Matrix is mentioned it will receive at least a 1.</p>
COMPONENT 5: EVALUATION METHODS – Weight 7,5%		
ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
EM1. Statement of the evaluation approach	<ol style="list-style-type: none"> 1. Approach used 2. Justification 	<p>Approaches: Utilization Focused Evaluation, Non-experimental, Random Control Trial, Program Theory, Contribution analysis, Outcome mapping, participatory evaluation, etc.</p> <p>It is possible that no methods will be mentioned; must be included in comments.</p>
EM2. Type of analysis considered	<ol style="list-style-type: none"> 1. Type of analysis 2. Reliability assurance 	<ol style="list-style-type: none"> 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.	<ol style="list-style-type: none"> 1. List of data collection methods and explanation 2. Description of different data sources 	<ol style="list-style-type: none"> 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. 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.
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.
COMPONENT 6: EVALUATION FINDINGS – Weight 25%		
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.”
COMPONENT 7: CONCLUSIONS – Weight 20%		

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
C1. Value	<ol style="list-style-type: none"> 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	<ol style="list-style-type: none"> 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

COMPONENT 8: RECOMMENDATIONS – Weight 20%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
R1. Clarity	<ol style="list-style-type: none"> 1. Are clear and concise 2. Are based on findings and/or conclusions of the report 	<ol style="list-style-type: none"> 1. (i.e. one or two sentences followed by explanatory text, if needed) 2. Directly linked to findings and conclusions
R2. Relevance	<ol style="list-style-type: none"> 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	<ol style="list-style-type: none"> 1. Specify who is called upon to act 2. Identify means for achievement 3. Specify priority or importance 	<ol style="list-style-type: none"> 1. Unit, department, office, etc. 2. Human, financial, material. 3. Low, medium, high.

COMPONENT 9: LESSONS LEARNED – Weight 2.5%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
LL1. Relevance	<ol style="list-style-type: none"> 1. Lessons learned are derived from findings 2. LL represent a relevant / new piece of information to be considered in the future. 	<p>The lessons deal with significant non-trivial matters. Summarize knowledge or understanding gained from experience related to the intervention under evaluation</p> <p>“Lessons learned are generalizations based on evaluation experiences that abstract from specific circumstances to broader situations”.</p> <p>Please notice LL are usually required in ToR but not addressed specifically in the report. Make a comment on that when this is identified.</p>

COMPONENT 10: GOOD PRACTICES – Weight 2.5%

ITEMS	ELEMENTS	RATING (0-5) & COMMENTS
GP1. Delimitation	<ol style="list-style-type: none"> 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	<p>1. The statements describe how the good practices should be implemented and by whom.</p> <p>2. The good practices explore the applicability in different contexts.</p>	Please notice GP are usually required in ToR but not addressed specifically in the report. Make a comment on that when this is identified.
OVERALL RATING		
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.

ANNEX XI – LIST OF INTERVIEWEES

NAME	POST	OFFICE	INTERVIEW METHOD
CASAGRANDE, Margaux	IDF	HQ	Skype interview
CASTELFRANCO, Alessia	Head of IDF	HQ	Skype interview
DE MULLER BARBAT Maria Teresa	ROMEO	Dakar	Skype interview
DUROCHER, Joannine	M&E Officer	Niger	Skype interview
EL MOAL, Abdel	M&E Officer	Senegal	Skype interview
FRANZETTI, Cristophe	Chief Evaluation officer	HQ	Skype interview
HARRIS Sarah Lynn	ROMEO	Vienna	Skype interview
KIM, Sokleang	ROMEO	Bangkok	Skype interview
MUTIE, Rogers	ROMEO	Nairobi	Skype interview
NISHIMURA, Eriko	Head of Office	Lesotho	Skype interview
REY, Jhonn	ROMEO	Buenos Aires	Skype interview
RIJKS, Barbara	Chief of Mission	Niger	Skype interview
SAAVEDRA, Silvia	M&E Officer	Colombia	Skype interview
TODOROVA, Irina	MPA thematic specialist	Cairo	Skype interview
VACA, Vanesa	ROMEO	San José	Skype interview
WAMBANDA, Angeline	ROMEO	Nairobi	Skype interview
ZECEVIC, Aida	Donor Relations Associate	Washington	Skype interview

ANNEX XII – DETAILED ANALYSIS OF REPORT QUALITY PER REGION

The following analysis put focus on the differences on the median rating per component and geographical area. As included in the main report, the following table shows the ranges in which the assessment is based.

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

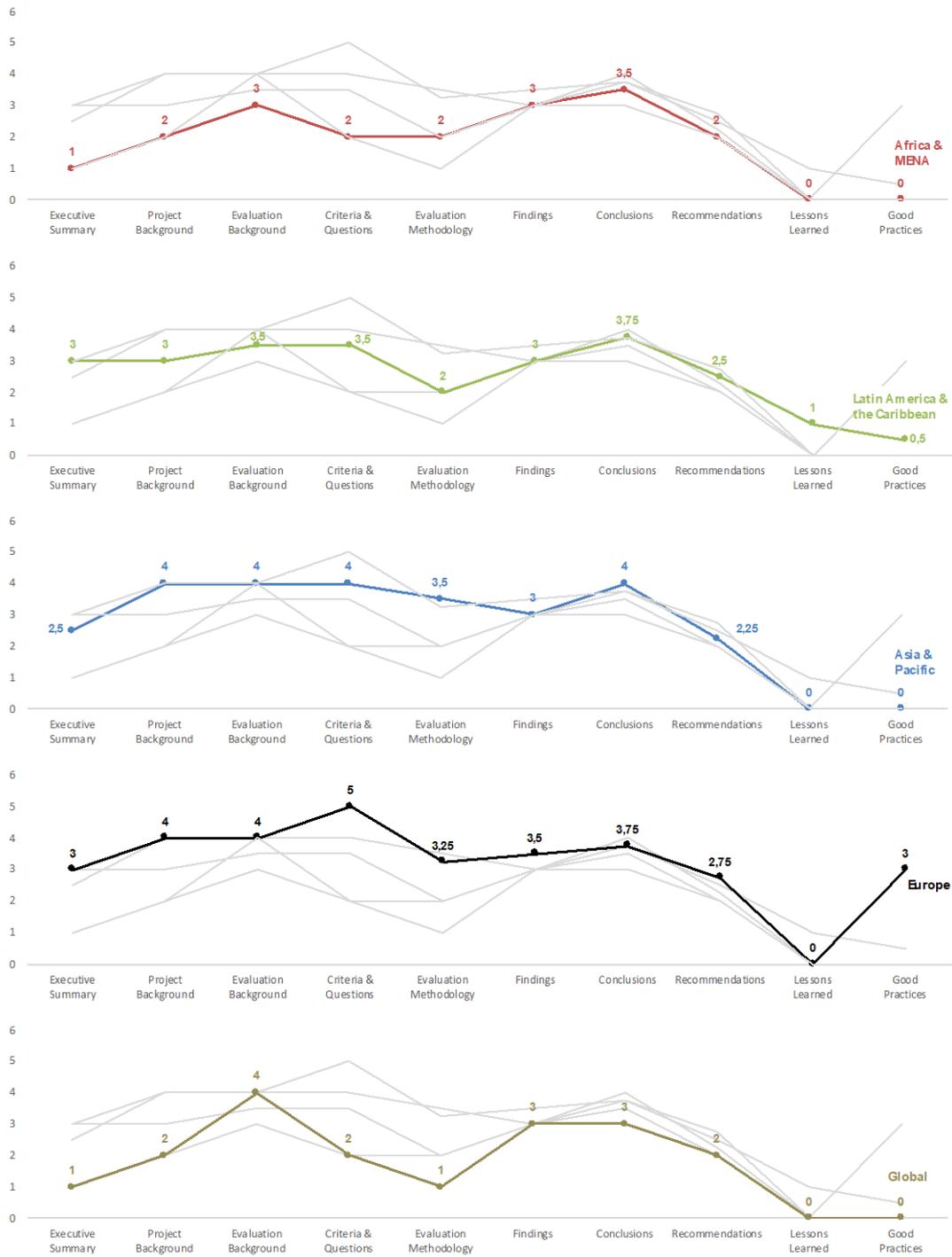
The result of the analysis for the component 1 (executive summary) reveals that only Latin America & the Caribbean (LATC) and Europe have a median rating that meets requirements (median rating of 3 or above). In terms of the quality of the project background section, reports from LATC have a median global rating that meets requirements, and median ratings of reports from Asia & Pacific and Europe exceeds requirements. All regions reached a median rating that meets requirements for the evaluation background section, which is in fact the component with the highest overall rating.

In terms of the quality of the project background section, reports from LATC have a median global rating that meets requirements, and median ratings of reports from Asia & Pacific and Europe exceeds requirements. All regions reached a median rating that meets requirements for the evaluation background section, which is in fact the component with the highest overall rating.

Concerning the component on evaluation criteria and questions, all regions reached a median rating that meets requirements, with the exception of reports from Africa and Global reports. The component on evaluation methodology is one with the lowest median overall ratings, and only reports from Asia & Pacific and Europe reached a median rating that meets requirements.

With regards to the component of evaluation findings, all regions reached a median rating that meets requirements. In this line, the component about evaluation conclusions is the second with the highest global median rating, and all the regions meets requirements. The overall median rating for reports from Asia and the Pacific exceeds requirements for the component about conclusions. On the other hand, none of the regions reached a median rating within the “meeting requirements” or “exceeding requirements” levels. Finally, and as already highlighted in the main report, components on good practices and lessons learned are the lowest rated components of them all. The median ratings for all the regions range from zero to one point, which reveals the limited attention given to those components in general.

The following figure reveals the median ratings per component and region. The ratings are expressed in the Y axis of the figure, and the components in the X axis.



ANNEX XIII – INCEPTION REPORT

FINAL INCEPTION REPORT

Meta-Evaluation of IOM's Internal and
External Evaluations (2017 – 2019)

January 2019

Consultancy: A Meta-Evaluation of IOM's Internal and External Evaluations

Timeframe of the interventions studied: January 2017 to December 2019

Locations (country, region, etc.) of the interventions under review: Worldwide

Consulting team: Salvador Bustamante (Coordinator), Leticia Bendelac, Jorge García and Juncal Gilsanz.

Organisation commissioning the consultancy: International Organization for Migration

Name of contact point for the consultancy: Diana Cartier

Consultancy budget: 29,650€

ACRONYMS

ALNAP-QP	ALNAP Quality Proforma
DAC	Development Assistance Committee
EPI	Evaluation Performance Indicator
GE	Gender Equality
GEROS	UNICEF Global Evaluation Report Oversight System
GEEW	Gender Equality and Empowerment of Women
GMDAC	Global Migration Data Analysis Centre
IOM	International Organization for Migration
M&E	Monitoring and Evaluation
MOPAN	Multilateral Organisation Performance Assessment Network
OECD	Organisation for Economic Co-operation and Development
OIG	Office of the Inspector General
UN-SWAP	United Nations System-Wide Action Plan
ToR	Terms of Reference
UN	United Nations
UNEG	United Nations Evaluation Group
USD	United States Dollar

BACKGROUND

Introduction

This report documents the results of the inception phase and activities directed to scope the meta-evaluation of International Organization for Migration's (IOM) during 2017-2019 (from now on, "the meta-evaluation").

The report provides information on the context within which the meta-evaluation is considered relevant. It also specifies the purpose of the objectives, scope and methodology, including the evaluation/research questions and evaluation methods, sampling strategy and the strategy for data collection, analysis and interpretation. The inception report together with the Terms of Reference (ToR) will guide the consultancy. The meta-evaluation will be conducted between November 2019 and March 2020 by an external consulting team (Artival Research & Evaluation). The IOM and other relevant partners will review submitted products, provide comments and sign-off the final report.

About the object of the meta-evaluation

As stated in the ToR, evaluation is an important learning and accountability tool on which there is an increasing emphasis placed both within the IOM and externally by donors and key stakeholders, encouraging the use of evaluation findings and recommendations in formulating policies and programmes.

To provide more quality assurance system, efforts are being made to continue building a culture of evaluative understanding and practice, particularly given the increased acknowledgment of the importance of evaluations for reporting on project results, lessons learning and identifying innovative and replicable practices. These efforts made by IOM are:

- Since late 2016, eight out of nine IOM regional offices have Regional Monitoring and Evaluation (M&E) Officers who provide guidance, capacity building and technical assistance to country offices in their respective regions;
- The Office of the Inspector General's Central Evaluation (OIG/Evaluation) approved the OIG Strategy for the Management of its Evaluation and Monitoring Functions 2018 – 2020;
- The OIG Evaluation team has grown to continue providing additional support to and continuity across the regions;
- In 2018, OIG also launched the IOM internal evaluator training aimed at contributing to the creation and development of a pool of internal evaluators able to carry out evaluations.

In this context, the meta-evaluation will assess the quality and use of internal and external evaluations (centralized and decentralized) conducted between 2017 and 2019 by IOM.

OBJECTIVES AND SCOPE

Objectives

In line with the ToR, the main objectives of the meta-evaluation are:

4. To assess the quality and use of internal and external evaluations (centralized and decentralized) conducted between 2017 and 2019, with the aim of providing actionable recommendations to enhance the quality and utilization of evaluations.
5. To provide IOM staff and partners with feedback on the quality of its evaluations, and if capacity-building efforts to support the conduct of evaluations have made a change in its quality.
6. To produce a full-fledged meta-evaluation tool in which to base the evolution of IOM's evaluation quality in the mid and long term.

The main users of this evaluation will be the OIG Evaluation staff and the Regional M&E Officers, whom together will be able to strengthen support and guidance in line with the findings of the report. In addition, this evaluation will be relevant for IOM's senior management as a mean of assessing how the Organization should improve the quality and use of its evaluation reports. Finally, IOM staff in general, and specifically programme managers and thematic specialists should take specific look at the quality and use of evaluations hence, this evaluation will provide better understanding of this aspect and how this can be improved.

Scope

In line with the ToR, the scope of the meta-evaluation will assess information contained in the internal and external evaluation reports (centralized and decentralized) for the period 2017 – 2019 and the ToRs (when available). This information will be complemented with the feedback provided by key stakeholder through interviews and an online survey. The total of internal and external evaluations to assess is 88. The Geographic coverage includes all the different regions and sub-regions where IOM operates. More information about the evaluation reports to be assessed can be found in section 3.4 (figure 1).

The **meta-evaluation will focus on maximizing the utility of its results and their actual use**. In order to do so, the meta-evaluation design and process will be carried out with careful consideration on meeting the information needs of primary and secondary audiences as tentatively expressed in table 1. In the interest of ensuring the optimal use of the findings, the meta-evaluation results should be presented constructively and appreciatively to the possible extent. This is key since the meta-evaluation standards set for the exercise go beyond the evaluation requirements in place during the period under evaluation. This is justified by the fact that these standards should become the basis for future meta-evaluation exercises, allowing evaluation quality comparisons over time.

Table 6. Intended users and expected use for the meta-evaluation

Type of User	Designation	Intended use
PRIMARY	OIG Evaluation staff Regional M&E Officers	<ul style="list-style-type: none"> ▪ Development of IOM's evaluation function. ▪ Improve preparation, implementation and resource management. ▪ Accountability ▪ Learning
	IOM's senior management and IOM in general	<ul style="list-style-type: none"> ▪ To improve the quality and use of its evaluation reports.

Type of User	Designation	Intended use
		<ul style="list-style-type: none"> To improve practices in the future as IOM staff manage, conduct or use evaluations
SECONDARY	Other UN agencies and organizations	<ul style="list-style-type: none"> Organizational learning/development
	Indirect beneficiaries / civil society	<ul style="list-style-type: none"> Learning/development, accountability

The work with primary intended users will not be limited to the provision of information; it will also include strategies and approaches to adequately implement and communicate findings, conclusions and recommendations. This analysis will also help the consulting team to formulate specific, realistic and relevant recommendations based on previous experiences, a participatory process with key stakeholders and a clear and logical, evidence-based connection among findings, conclusions and recommendations.

APPROACH AND METHODOLOGY

The purpose of this section is to establish the approach and methodology that will guide the data collection, analysis, interpretation and the way in which the final deliverables will be framed.

Approach

The driving questions for the meta-evaluation can be summarized as “What is the quality of the evaluations conducted and what is required to strengthen IOM’s evaluation function in the future?” and “How useful were the evaluations conducted for the IOM in the proposed period?” The following table reveals the key concerns under consideration and the different areas considered for the analysis.

Table 7. Key areas of analysis of the meta-evaluation

Key concern	Areas of analysis
Evaluation quality	<ul style="list-style-type: none"> Evaluation and ToR quality based on reference documents, and including cross-cutting issues 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 on the quality of evaluation through time
Evaluation capacities	<ul style="list-style-type: none"> Quality per type of evaluator/agent
Evaluation use	<ul style="list-style-type: none"> Strategies in place toward maximizing evaluation use. Issues that facilitated or hindered evaluation use

Methodology

The meta-evaluation will follow an utilization focused approach³³ and will be developed through a mixed-method concurrent triangulation³⁴ research strategy. The consulting team will collect both quantitative and qualitative data concurrently and then compare databases to

³³ https://www.betterevaluation.org/en/plan/approach/utilization_focused_evaluation

³⁴ Creswell, J. (2003) “Research design”, third edition.

establish convergences, divergences or a combination of both. Therefore, a process of confirmation, disconfirmation, cross-validation and corroboration will be carried out in order to obtain key findings in which to base conclusions and recommendations. In practice, this means that quantitative data will provide findings to trace statistical trends, whilst qualitative data obtained during the assessment process will provide explanations and nuances of the quantitative trends found. The strategy will be conducted in an iterative way to ensure the highest methodological robustness.

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

Table 8. Methodological approach

Methodological strategies	Research Techniques	Data gathering tools
<ul style="list-style-type: none"> ▪ Qualitative strategies (Discourse analysis) 	<ul style="list-style-type: none"> ▪ Semi-structured interviews ▪ Evaluators comments and perceptions 	<ul style="list-style-type: none"> ▪ Interview protocols ▪ Scoring tool
<ul style="list-style-type: none"> ▪ Quantitative strategies (Descriptive and inferential statistical analysis) 	<ul style="list-style-type: none"> ▪ Analytical framework ▪ Online survey 	<ul style="list-style-type: none"> ▪ 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, and data gathered through qualitative methods (interviews, document review and the scoring tool) 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 will use different sources of information in order to fill the gap between the absence of information in some cases, non-updated information or unreliable information. Data obtain from the assessment of reports and ToR 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 three complementary research techniques considered are:

4. *Assessment with the scoring tool:* The analytical framework and its scoring tool will provide the quantitative data in which to base the meta-evaluation results, allowing the emergence of statistical trends by means of proportions, measures of central tendency, and statistical association through correlation coefficients. The statistical association analysis will be used to determine the extent to which evaluation budgets and the type of agent conducting the evaluation (internal/external) are linked to the ratings given in terms of evaluation quality. Once the quantitative database will be ready, specific software will be used to conduct the analysis and produce the visuals (SPSS, Tableau and Excel). The tool will also gather qualitative information about perceptions of the evaluators that will go beyond the rating given. This will be done through the comments made to each particular rating.

5. *Semi-structured interview*: interviews with a select number of IOM staff will be the other crucial method in the meta-evaluation, as they will allow to triangulate information, point of views and opinions of different stakeholders. The views and information provided will deepen the findings of the evaluation. Around 10 interviews are expected to take place, targeting regional monitoring and evaluation officers as well as a selection of key informants suggested by OIG and the reference group.
6. *Survey*: an on-line survey was developed and will be launched to complement qualitative information with related quantitative data. The questionnaires will be sent to both internal and external evaluator, as well as to the managers of the evaluations under review.

The consulting team will develop the tools concurrently, and a first draft of KII protocol and survey questionnaires are attached for reference (annex 4 & 5). These are informed by the key meta-evaluation questions considered, as well as by the key areas under assessment included in the Scoring tool.

Table 9. Meta-evaluation questions and techniques

Meta-evaluation questions	Concept	Research Tools
9. To what extent are IOM evaluation reports and their ToRs meeting the UNEG evaluation standards using evaluation criteria and including crosscutting issues?	Quality of evaluations	Scoring tool Interview Survey
10. 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
11. Are evaluation reports and TORs developed based on the recommended templates provided by IOM guidelines?	Quality of evaluations	Scoring tool
12. Have the time and cost of/resources assigned to the evaluations influenced the quality of the final product?	Quality of evaluations	Scoring tool Interviews
13. What can be determined about the overall quality of the evaluation reports and where do the greatest opportunities for improvement lie?	Quality of evaluations	Scoring tool Interview Survey
14. During the period covered, are there any changes in the quality of the evaluations conducted and if yes, what factors may have contributed towards these changes?	Quality of evaluations	Scoring tool Interviews
15. What patterns of quality exist (if any) between those conducted by the trained internal evaluators, non-trained internal evaluators, and external evaluators? What strengths and gaps in evaluation competencies do the processes and reports indicate?	Evaluation capacities	Scoring tool Interview Survey
16. How were/are evaluations used and promoted? What factors facilitated or hindered the use of the evaluations?	Use of evaluations	Interview Survey

As revealed in table 4, the scoring tool specifically designed for this exercise will play a major role in the analysis. The tool is based on the evaluation questions considered and a series of rubrics and ratings specifically developed to address and judge the extent to which each

particular 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 four categories as described in table 5.
- **Components:** second level of analysis, there are the key sections in which the assessment of each dimension is based.
- **Items:** represent the general criteria consider to adequately assess the compliance with each of the different components; they indicate key concepts to be considered under each specific component.
- **Elements:** represent the ultimate sub-criteria in which the assessment is based. One item can comprise one or more elements, and the assessment is based on the compliance with the set of elements consider for each item.

Table 10. Dimensions under analysis through the meta-evaluation tool

DIMENSION	UTILITY
4. Quality of evaluation reports	<ul style="list-style-type: none"> ▪ Collect information about the quality of evaluation reports. ▪ Rate the quality and relevance of the information within reports.
5. Comprehensiveness of evaluation report	<ul style="list-style-type: none"> ▪ Summary and checklist of the information contained in the evaluation reports.
6. Comprehensiveness of ToR	<ul style="list-style-type: none"> ▪ Assess the key sections and information contained in the terms of reference proposed by IOM to evaluators.

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

Table 11. Items and elements consider for component "Recommendations" in Dimension 1

COMPONENT 8: RECOMMENDATIONS – Weight 20%	
ITEMS	ELEMENTS
R1. Clarity	<ol style="list-style-type: none"> 1. Are clear and concise 2. Are based on findings and conclusions of the report
R2. Relevance	<ol style="list-style-type: none"> 1. Address key issues and are useful, tied to the object and purposes of the evaluation. 2. Address cross-cutting issues
R3. Responsibility	<ol style="list-style-type: none"> 1. Specify who is called upon to act 2. Identify means for achievement 3. Specify priority or importance
R4. Actionability	<ol style="list-style-type: none"> 1. Indicate the specific courses of action needed to remedy/or continue with the current situation 2. Recommend a time frame

Once all elements under a particular component are considered, a series or rubrics will be used to determine the level of compliance with established criteria. See next section for more details.

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

- A six-point scale included for Dimension 1 on quality of evaluation reports, which will represent the main scoring system. The different rubrics are expressed in table 7. This is an ordinal scale and therefore the analysis would take into consideration frequencies, proportions, mode and medians, but not means. **Finally, a global rating will be provided to each report** considering the ratings given and the relative weight consider for each of the components (see figure 1).
- A dichotomous scale included for Dimension 2 and 3 (comprehensiveness of reports and ToR). This is a nominal scale and therefore the analysis will consider mainly frequencies and proportions.

Table 12. 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 of the elements are present, and at least one of the elements present is incomplete	Somewhat unsatisfactory Not all of the elements are present, but those present are complete.	Somewhat satisfactory All of the elements are present but not all of them are complete	Satisfactory All of 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 the elaboration of 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 coded into different key 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 18- Relative weight for the rating of each component

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

In terms of enhancing the validity of the evaluation, the combination of proposed methods will allow the triangulation of information, strengthening the validity of the study. The evaluation 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

researches or evaluators). Measurement validity (the extent to which evaluation indicators capture the concept of interest) will be revised and presented for validation to key evaluation audiences.

In order to ensure the reliability of data-gathering tools, the evaluation team will conduct several reliability procedures. The scoring tool and the different rubrics will be piloted to test their relevance and fit for purpose. All team members will assess the same evaluation report in order to:

1. Compare the scores given by each team member and the reasoned exposition of each component to ensure the identification of discrepancies in the interpretation of criteria and rubrics. This will ensure the establishment of a common system of rating around the elements, items and components, and therefore the inter-rater reliability. 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.
2. Determine the level of effort required for the assessment of an evaluation to establish the final number of elements under analysis, and once confirmed by IOM, the final number of evaluation reports.

The contrast between the scores given by each team member and the reasoned exposition of each item will ensure the identification of discrepancies in the interpretation of criteria and scales and the agreement upon a common system of evaluation guidelines around the established rubrics. This revision will spark a methodological discussion about the boundaries and shared definitions of the concepts provided 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 the team members understand each key concept similarly. This will be accomplished by defining each code clearly; regularly comparing data with codes and by crosschecking codes developed by different team members. In the third phase of the meta-evaluation the draft report will be circulated to the different stakeholders, which will not only allow each partner to be involved in the evaluation, but will improve the initial draft overall: by completing, adjusting or qualifying the whole evaluation as well as deepening the lessons learned and the recommendations.

Two evaluation reports (one internal and one external) were piloted. The external evaluation report piloted is the “2014 - 2018 Final Evaluation IOM X Asia Campaign” (Num. 69 of Annex 3). This evaluation has been chosen because it is an external final evaluation of a regional program drafted in 2018 with several documents (ToR, evaluation matrix, etc.) that will facilitate the analysis. The internal evaluation piloted is a mid-term evaluation conducted for the project titled “Strengthening migration management in Niger” conducted in 2019.

As a result of the pilot exercise, the meta-evaluation rubrics were modified to ensure higher levels of interrater reliability (see table 7), as well as key elements presented in the meta-evaluation tool (annex 1, 2 and 3).

Once the piloting process is over, and the analysis of all the evaluation reports takes place, two will be the key strategies considered to ensure interrater agreement and interrater consistency

(both key components of interrater reliability). During the start of the data collection process, another round of verification will take place in order to ensure that the understanding of components, items and elements of the meta-evaluation tool remains the same across the coders. Additionally, the meta-evaluation report will specifically 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) 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, KII and online survey designed (both quantitative and qualitative). Secondary data is considered all the information in which the rating and analysis is based, as described in table 8.

Table 13. Secondary data sources

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 (UN SWAP EPI, MOPAM, etc.)	Contextualization and deep understanding of key frameworks and strategies considered within IOM.
The evaluation reports and ToRs	Key information in which to base the assessment and the meta evaluation

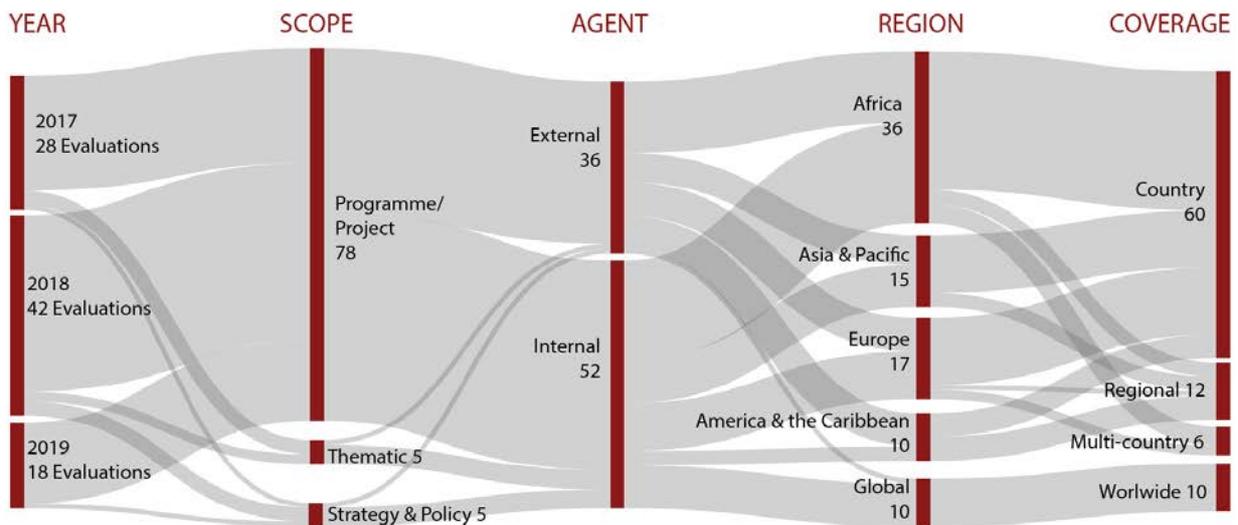
Sampling procedures

A total of 88 evaluation reports will be included in the assessment. Some of the particularities of these reports (year, scope, etc.) are revealed in figure 2. The entire universe of evaluations conducted during the assessment period is included in the assessment. Therefore, no sampling strategies were needed. With regards to the key informant for the interviews, the sample will follow a purposeful sample approach, selecting those evaluation managers at the regional level and OIG representatives. Finally, the survey will target every manager involved in evaluations during the period considered (2017-2019), as well as internal and external evaluators.

Cross-cutting issues

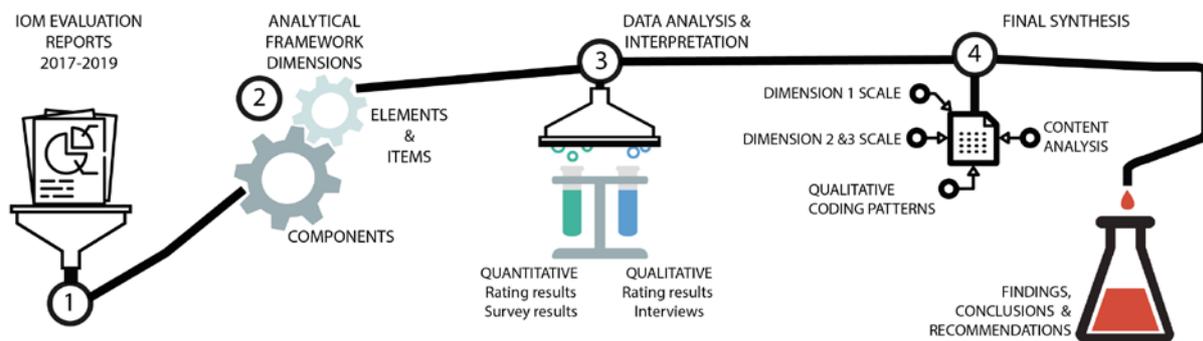
The crosscutting issues considered for the assessment are gender, rights-based approach, and environmental sensitivity and sustainability. The assessment will determine the extent to which these issues are taken into consideration when designing evaluations, as well as when defining findings, conclusion and recommendations.

Figure 19. Overview of evaluations under assessment



The following figure represents the **overall process for the assessment**, from the final universe of evaluation report under consideration to the findings, conclusions and recommendations of the final report.

Figure 20. Overall process of the meta-evaluation



Norms and standards

The consulting team will ensure that the meta-evaluation process respects the UNEG basic premises of ethical and professional behaviour. **The meta-evaluation will be conducted following relevant IOM policies, as well as UNEG norms and standards³⁵.** The consultants will be fair and carry out the assessment with integrity and honesty. The consultants declare not to have any conflicts of interest, as they have not directly or indirectly participated in any of the interventions under review, nor have a personal relationship with managers and/or consultants involved in those.

LIMITATIONS OF THE META-EVALUATION

³⁵ Click to access the UNEG standards.

The assessment team has identified the following risks and limitation for the meta-evaluation:

Table 14. potential risks 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 by internal evaluator	The team will include and 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 o externally. This could lead to substantial differences within the reports that could affect their quality.	The meta evaluation will take into account this dual nature and will considerer any difference concerning to this typology.
Large range of migration thematic areas and sectors	The evaluations 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 take into account this and will considerer any difference concerning to this typology.
Incomplete datasets	The consulting team will require reviewing documentation including ToR, evaluation reports and other relevant materials (resources and budgets). However, there might be limited information for some of the projects/programs involved.	The consulting team will inform about the missing data or documentation required.
Interviews and surveys limitations	Some of the evaluations were carried out by externals, others by staff who are no longer in the organizations, so the access to them might not be possible.	In case this limitation arises, the evaluation team will try to contact those staff members or external teams who have participated 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 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 manager to

RISK/LIMITATION	EXPLANATION	MITIGATING STRATEGY
		reschedule key milestones as required. This is reflected for the Inception phase in the updated timeframe.

SCHEDULE AND DELIVERABLES

Timeframe and work plan

The meta-evaluation will be implemented from the 25th of November 2019 until 6th March 2020. **The consulting team will be engaged for a total of 50 working days.**

Table 15. Key deliverables and milestones

Tasks	Timeframe	Product
Contract formalization	25 th of November 2019	Contract
Initial briefing with the meta-evaluation manager	28 th of November 2019	Initial notes
Refinement of methods, tools and questions	10 st of December 2019	Draft inception report
IOM comments on the draft inception report	18 th December 2019	Comments
Pilot of the analytical framework and results	7 th January 2020	Final Inception report with result of pilot
Validation meeting	9 th January 2020	
Data screening and synthesis	30 th January 2020	-
Draft final report	25 ^h February 2020	Draft final report
IOM comments on the draft final report	28 th February 2020	Comments
Final report and presentation	6 nd March 2020	Final report

Deliverables

- **An inception report** including a draft report outline, a list of possible additional data/information sources to be consulted; and a detailed methodology and 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); the methodological strategy; possible inclusion/exclusion criteria for the selection of documents; the means to address risk of bias and quality control. The approach taken should take into account the specificities of evaluation reviews of international development and humanitarian action interventions and

should consider criteria, questions, and methods suggested in guidance on evaluating institutional gender mainstreaming developed by UNEG dated April 2018. OIG and the evaluation staff will be invited to review and comment on the inception report, and the consultants will respond in writing to all received comments. The inception report must be approved by IOM before work can proceed.

- **A draft final report** based on the approved inception report. The draft report will present the analysis of the selected documents and key findings, and will include an executive summary. It will be written in English and be no longer than 50 pages, including the executive summary, but 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. OIM will conduct quality assurance/peer review of the draft version of the report.
- **A consolidated final report**, which will present the completed analysis and must respond appropriately to comments and feedback from key IOM officials. The final report must have a high-quality executive summary no longer than 10 pages.