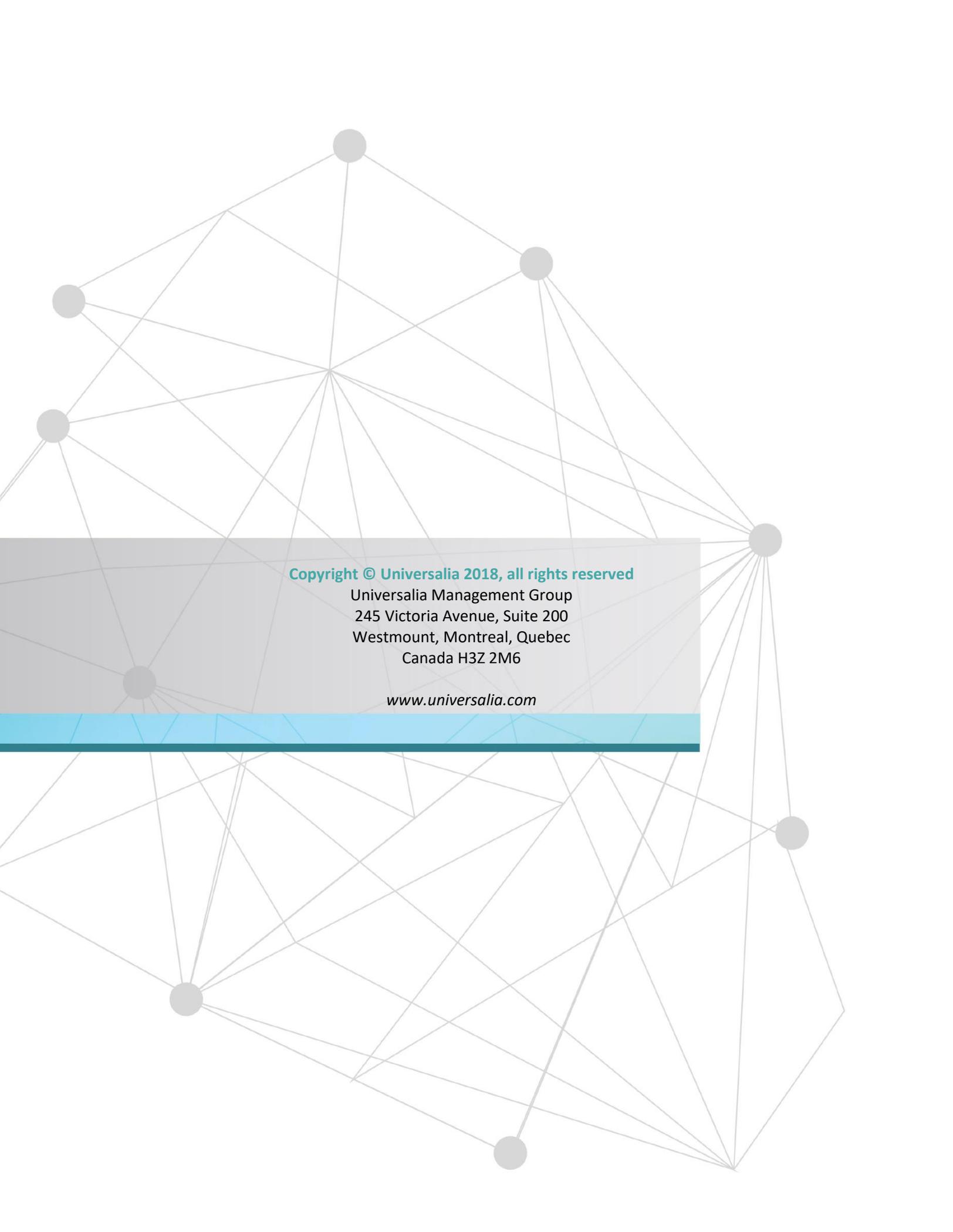


Quality Assessment of a Sample of Evaluation Products

Commissioned by the African Development Bank

FINAL REPORT | MAY 2018



Copyright © Universalialia 2018, all rights reserved

Universalialia Management Group
245 Victoria Avenue, Suite 200
Westmount, Montreal, Quebec
Canada H3Z 2M6

www.universalialia.com

Acknowledgments

Universalialia would like to thank the IDEV team in Abidjan for their cooperation and assistance in facilitating our work on this assignment. We gratefully acknowledge the valuable technical observations and logistical support provided by IDEV throughout the quality assessment process, as well as the documentation and feedback on deliverables provided to the quality assessment team in a timely manner.

Acronyms

AfDB	African Development Bank
ECG	Evaluation Cooperation Group
GE	Gender Equality
IDEV	Independent Development Evaluation
OECD-DAC	Organisation for Economic Co-operation and Development/Development Assistance Committee
QA	Quality Assessment
RBM	Results-Based Management
ToC	Theory of Change
ToR	Terms of Reference
UMG	Universal Management Group

Executive Summary

Quality Assessment Purpose, Objectives, and Scope

This quality assessment (QA) of a sample of IDEV evaluation products sought to enhance learning that will help IDEV continue to improve the quality of its evaluations. It sought in particular to determine if there has been improvement in the quality of evaluations that IDEV has delivered since the beginning of the implementation of its 2013-17 strategy by comparing the quality of evaluations delivered before the implementation period of the strategy (i.e. 2009-2012) against evaluations delivered towards the end of its implementation (i.e. 2015-2017).

A total of 22 evaluations were sampled, representing almost 50% of the 47 total eligible evaluations. The sample is representative of the main types of evaluations conducted by IDEV which were matched by type insofar as possible between each period under review: country evaluations, thematic and sector evaluations, corporate and process evaluations, and others such as impact evaluations, regional evaluations, cluster evaluations, and evaluation syntheses. Nine evaluations were selected from the first period and 13 from the second period.

Methodology

This quality assessment assignment involved a meta-analysis of various evaluation products. The assessment was guided by OECD-DAC quality standards for evaluation and the Evaluation Cooperation Group's good practice standards for evaluation. It essentially comprised a desk review of the sample of 22 evaluations, including summary reports and corresponding documents such as full technical reports, ToR, approach papers, inception reports, and any relevant documentation available. Quality assessment findings and recommendations were presented to IDEV personnel and further refined during an interactive workshop held in Abidjan on March 27, 2018.

A standard template including 26 QA criteria organized in four clusters, each in accordance with international standards, was used to assess and rate the criteria based on a 4-point scale. The clusters comprise: context, object, purpose, objectives, and scope of the evaluation (cluster 1); design and methodology (cluster 2); process (cluster 3); and clarity and robustness of reporting (cluster 4). Average cluster ratings were calculated and colour-coded and overall evaluation assessments were determined based on the percentage of individual criteria ratings that were rated "satisfactory" or "highly satisfactory" (3 and 4 scores) only.

The QA process involved three steps: individual assessment of each evaluation by two assessors; consolidation into a single assessment for each evaluation; and harmonization of all assessments by the full three-member team to ensure overall consistency and quality. Quantitative and qualitative analysis techniques were used to inform the findings of the QA.

Some QA limitations included: limited data available for some evaluations; limited statistical significance due to the small sample size; and limited comparability between the first and second periods for some types of evaluations in the sample that were not produced in the first period.

Key Findings

For all evaluations combined, **the majority of quality assessment criteria were rated “satisfactory” and “highly satisfactory”**, with a fairly even distribution of these ratings among the four clusters of criteria.

The most frequent rating for all four clusters of criteria was “satisfactory” with design and methodology showing the greatest variation in ratings among all evaluations.

There is a marked improvement in the quality of evaluations over time from 2009-12 to 2015-17, with all 13 second period evaluations receiving either a “satisfactory” or “highly satisfactory” average rating for all clusters. Among the four clusters, cluster 2 on design and methodology showed the most improvement.

Among the seven criteria related to the context, object, purpose, objectives, and scope of the evaluation, description of the object of the evaluation and scope of the evaluation are the highest rated criteria, and cross-cutting issues the lowest rated criterion, in both periods under review.

Within the design and methodology cluster, data collection, analysis and sampling is the highest rated criterion, with no evaluations showing a “highly unsatisfactory” rating on this criterion, **and alignment of evaluation questions with criteria and indicators in an evaluation matrix is the lowest rated criterion**, although there was significant improvement on this criterion in the second period evaluations.

The rating scale used by IDEV evaluations was also found to be problematic. The assessment team acknowledges IDEV’s recent decision to replace the cumbersome 6-point rating scale with a 4-point scale for future evaluations. However, the labels on the new scale – “highly unsatisfactory”, “unsatisfactory”, “satisfactory”, “highly satisfactory” – remain problematic to a certain extent as they do not allow for a more nuanced assessment between “unsatisfactory” and “satisfactory” where this would be relevant or applicable.

The links between evaluation products for each evaluation were logical overall, despite some overlap in information in some cases or sometimes key elements missing from the summary report likely due to restrictions in the number of pages allowed for these reports.

Evaluation principles adhered to by IDEV – independence, impartiality, ethics, coordination and alignment, partnership and capacity development – , while they may have effectively been followed, were not satisfactorily documented across both periods. A process-related criterion that was rated somewhat lower in the second period concerns stakeholder engagement: key evaluation stakeholders and their roles and responsibilities were not always systematically identified.

The large majority of evaluation reports were well structured and included a concise executive summary, **presented evidence-based findings, addressed the evaluation criteria and questions, and included relevant recommendations but did not always explicitly identify the sources from which data were derived.**

All evaluations demonstrated clear linkages between the evidence, findings, conclusions, lessons learned, and recommendations overall, despite the lack of distinction sometimes between the elements that each should address. Cross-cutting issues were not well reflected in the findings and conclusions in almost half the reports reviewed spanning both periods while lessons learned, where applicable, were either not included or not correctly worded as lessons in half of the reports, particularly those from the second period. Recommendations, while relevant overall, often did not include a timeline for implementation.

Recommendations

1. Consider adding a requirement for stakeholder mapping. This requirement would be stipulated in the initial evaluation planning documents and the stakeholder map developed by the evaluators and presented in both the inception and evaluation technical reports (in the annexes if too detailed). This would ensure all key stakeholders (internal and external) are identified and their roles and contributions clearly described.
2. IDEV should define what cross-cutting themes to prioritize on the basis of what AfDB policy has identified as important and should identify which of these themes are relevant for each type of evaluation:
 - Relevant cross-cutting themes should be explicitly identified at the outset (inception phase) and effectively integrated in a cross-cutting manner, beginning with the description of the context and object of the evaluation through the evaluation objectives, scope, findings, conclusions, and recommendations;
 - Where the absence of disaggregated data on these themes impedes their full assessment, the evaluation can be an opportunity to probe these issues further through their integration in the evaluation framework and corresponding primary data collection instruments and to recommend to management that more attention be paid to them in the monitoring of Bank operations.
 - With respect to gender equality in particular, while it may not be deemed a relevant cross-cutting theme in all types of evaluations conducted by IDEV, given that the AfDB, like all international financial institutions, has a policy to promote gender equality in development work, IDEV should attempt to systematically integrate gender considerations into all evaluations that involve the assessment of projects as well as other types of evaluations where it would be relevant to do so.
 - Where support is needed for the effective integration of these issues, IDEV should seek expert technical assistance through specialized workshops, webinars, etc. aimed at building the capacities of task managers in these areas.
3. Ensure that OECD-DAC and AfDB-IDEV evaluation principles related to evaluation independence, impartiality, ethics, partnership, coordination, and capacity building are clearly and explicitly stated and explained in the various evaluation products, from the design to final report stage, and considered as part and parcel of the evaluative process. Evidence of consideration of these principles could be provided following the description of the evaluation methodology, in the

introduction of various data collection tools, and in a disclaimer in the opening pages of evaluation reports. While a disclaimer could be included in both the summary and technical reports, further information on evaluation principles could be included in the methodology annex of summary reports given restrictions on length of these reports.

4. Consider re-labeling the 4-point rating scale to allow for a perspective that more positively reflects some degree of progress made toward satisfactory performance. We recommend replacing the existing rating labels with the following: “unsatisfactory”, “partly unsatisfactory”, “satisfactory” and “highly satisfactory”.
5. Clearly identify the basic elements that each evaluation product should contain and clarify the circumstances under which ToR, approach papers, and inception reports should be produced.
6. Develop guidance for evaluators with clear definitions of key evaluation terms to ensure a distinction is made between the type of information that should be included in the different sections of the evaluation report, e.g. evaluation findings, conclusions, lessons learned, and recommendations.
7. Ensure that recommendations are discussed systematically with the relevant operations sections of the Bank and clearly prioritized or classified with a corresponding timeframe to help inform decision making. While avoiding the formulation of recommendations that are too prescriptive, IDEV should identify issues of concern within the recommendations to connect them more specifically to the evaluation findings reported.
8. Ensure the annexes in the inception report include the data collection tools and the annexes in the summary report contain an abridged evaluation matrix, a key evaluation tool that links the scope of the evaluation to the evaluation methodology. The abridged matrix could include the evaluation criteria and main corresponding questions as well as a sample of key indicators – both quantitative and qualitative where relevant – and the data collection methods for each main question.

Contents

EXECUTIVE SUMMARY	V
1 INTRODUCTION AND CONTEXT	1
1.1 Report Contents.....	1
1.2 AfDB-Independent Evaluation (IDEV) Strategy 2013-2017	1
2 PURPOSE, OBJECTIVES, AND SCOPE OF THE ASSIGNMENT.....	2
3 METHODOLOGY.....	3
3.1 Meta-Analysis Design and Data Collection.....	3
3.2 Sampling, Template, and Analysis Methods.....	4
3.2.1 Sampling.....	4
3.2.2 Assessment Template and Criteria.....	5
3.2.3 Rating Scale and Rationale for Ratings.....	6
3.2.4 Quantitative and Qualitative Analysis.....	10
3.3 Limitations and Mitigating Actions.....	10
4 MAIN FINDINGS	12
4.1 Overall Analysis of Quality of Evaluations	12
4.1.1 Overall Ratings by Cluster	12
4.1.2 Overall Comparison across Types of Evaluations between First and Second Periods.....	13
4.2 Comparative Analysis of Strengths and Weaknesses of Evaluations by Criteria between First and Second Periods	16
4.2.1 Context, Object, Purpose, Objectives, and Scope of Evaluations	16
4.2.2 Design and Methodology	18
4.2.3 Process	20
4.2.4 Clarity and Robustness of Reporting.....	23
5 CONCLUSIONS, LESSONS LEARNED, AND RECOMMENDATIONS.....	26
5.1 Conclusions.....	26
5.2 Lessons Learned.....	27
5.3 Recommendations.....	27

Figures

Figure 3.1	Distribution of evaluations by type and evaluation products (n=22)	5
Figure 3.2	Three-step quality assessment process	9
Figure 3.3	Distribution of evaluations for among assessors	9
Figure 4.1	Percentage of criteria rated 3 and 4 by cluster, all evaluations.....	12
Figure 4.2	Average cluster ratings, all evaluations (n=22)	13
Figure 4.3	Comparison of percentage of criteria rated 3 and 4 by evaluation period and type.....	14
Figure 4.4	Comparison of average cluster ratings between first period and second period, all evaluations (first period n=9, second period n=13)	14
Figure 4.5	Comparison average cluster ratings by type of evaluation, first period and second period ..	15
Figure 4.6	Cluster 1 Ratings: Comparison between Periods (first period n=9, second period n=13)	17
Figure 4.7	Cluster 2 Ratings: Comparison between Periods (first period n=9, second period n=13)	20
Figure 4.8	Cluster 3 Ratings: Comparison between Periods (first period n=9, second period n=13)	22
Figure 4.9	Cluster 4 Ratings: Comparison between Periods (first period n=9, second period n=13)	24

Tables

Table 3.1	Rating Scale	6
Table 3.2	Case examples of four different ratings and assessments for one criterion	7

Appendices

Appendix I	List of Evaluations Reviewed.....	31
Appendix II	Quality Assessment Template.....	33
Appendix III	Percentage of criteria rated Satisfactory and Highly Satisfactory, All Evaluations	41
Appendix IV	Quality Assessment Results by criteria, All Evaluations	43
Appendix V	Quality Assessment Results by criteria, First Period Evaluations (2009-2012)	45
Appendix VI	Quality Assessment Results by criteria, Second Period Evaluations (2015-2017)	47
Appendix VII	Quality Assessment Results by criteria, Country evaluations	49
Appendix VIII	Quality Assessment Results by criteria, Corporate Evaluations.....	51
Appendix IX	Quality Assessment Results by criteria, Thematic and Sector Evaluations	53
Appendix X	Quality Assessment Results by Criteria, Other Evaluations.....	55
Appendix XI	Quality Assessment Results by clusters of criteria, All Evaluations.....	57

Appendix XII Quality Assessment Results by cluster of criteria, Country Evaluations	59
Appendix XIII Quality Assessment Results by cluster of criteria, Corporate Evaluations	61
Appendix XIV Quality Assessment Results by cluster of criteria, Thematic/Sector Evaluations	63
Appendix XV Quality Assessment Results by cluster of criteria, other Evaluations	65
Appendix XVI Quality Assessment Results by cluster of criteria and type of evaluation	67
Appendix XVII Quality Assessment Results by cluster of criteria, evaluation periods, and types of evaluation	69
Appendix XVIII List of Documents Consulted	71
Appendix XIX Abidjan Workshop Agenda and Report	81

1 Introduction and Context

1.1 Report Contents

The Universalia Management Group Limited (henceforth, “Universalia”) is pleased to submit this final synthesis report to the African Development Bank (AfDB)-Independent Development Evaluation (IDEV) of the quality assessment (QA) of a sample of its evaluation products.

Included in this introduction is an overview of the context of this assignment beginning with a brief description of the AfDB-IDEV Strategy 2013-17 and of an assessment of the implementation of the Strategy currently underway of which this QA is a part. Following this introduction, the report includes four main sections:

- Section 2 provides an overview of the purpose, objectives, and scope of the assignment.
- Section 3 presents the methodology, including the design, data collection and sampling methods, description of QA template, analysis methods, limitations and mitigating actions.
- Section 4 presents the main findings of the assessment, beginning with an overall analysis of the quality of evaluations, followed by a comparative analysis of the strengths and weaknesses of the evaluations by cluster and criteria across evaluation types for each period under review (2009-12 and 2015-17).
- Section 5 presents the quality assessment conclusions, lessons learned, and recommendations.

Some case examples are highlighted in textboxes in the findings section. Appended to the report are the: list of evaluations reviewed; quality assessment template; detailed data figures; list of documents consulted for this assignment; and the QA workshop agenda and report, including list of participants.

1.2 AfDB-Independent Evaluation (IDEV) Strategy 2013-2017

The Bank’s first independent evaluation strategy was developed to effectively link the Bank’s Independent Evaluation Policy with IDEV programming over a five-year period between 2013 and 2017. The policy “identifies three complementary objectives for independent evaluation: learning, accountability, and promoting an evaluation culture” with the aim of supporting the improvement of development effectiveness.¹ In pursuit of these three objectives, IDEV has re-examined its mix of evaluation products, increasingly diversifying its portfolio to include more cluster, country strategy, and corporate evaluations as well as evaluation syntheses while introducing new products such as regional

¹ IDEV, *African Development Bank Independent Evaluation Strategy 2013-2017*, February 2013, p. 17.

integration strategy evaluations and impact evaluations. It has maintained its traditional output of thematic and sector evaluations but reduced the number of project performance evaluations.²

This quality assessment of IDEV's evaluation products carried out by Universalialia is part of a comprehensive assessment of IDEV's implementation of the strategy that is currently being conducted internally by IDEV (i.e. through self-assessment). IDEV commissioned an external firm to conduct the quality assessment to enhance transparency, avoid potential conflict of interest, and bring additional expertise to the exercise. The results of this assessment feed into the broader assessment of the independent evaluation strategy implementation, and inform new strategic directions to the AfDB independent evaluation function, by either replacing or extending the strategy in 2018.

2 Purpose, Objectives, and Scope of the Assignment

While this quality assessment will provide accountability regarding the improvement of evaluation quality, its main purpose was to enhance learning that will help IDEV continue to improve the quality of its evaluations.

This quality assessment in particular sought to determine if there has been improvement in the quality of evaluations that IDEV has delivered since the beginning of the implementation of the strategy. Therefore, the assessment aimed at comparing the quality of evaluations delivered before the implementation period of the strategy (i.e. 2009-2012) with evaluations delivered towards the end of its implementation (i.e. 2015-2017).

The scope of this assignment as defined in the ToR and Universalialia's technical proposal was further refined during the inception phase. It is circumscribed by the following key elements:

- **Object of the quality assessment:** As agreed with IDEV, the total sample includes 22 evaluations (see Appendix I for the full list of the evaluations reviewed), which represents almost 50% of the total evaluation population of 47 (see total evaluation population below). The assessment did not include joint evaluations that were not led by IDEV and did not cover other related evaluation products such as briefs, magazines and events.
- **Typology of evaluations:** The sample is representative of the main types of evaluations conducted by IDEV as defined in its strategy and inception report for assessing the strategy's implementation³, i.e. country strategy evaluations, thematic and sector evaluations, corporate and process evaluations, and others such as impact evaluations, regional evaluations, cluster evaluations, and evaluation syntheses. It did not include individual project performance evaluation reports (PPERs).

² Ibid., p. 19. While the output of project evaluations has decreased, it should be noted that each cluster evaluation includes assessment of several projects.

³ IDEV, *Assessing implementation of the AfDB Independent Evaluation Strategy 2013-17*, Inception Report V3, November 2017.

- **Period under review:** As per the QA ToR and inception report for assessing IDEV's strategy, the sample of evaluations conducted before the beginning of the strategy's implementation (2009-12) are referred to as first period evaluations while evaluations completed toward the end of the strategy (2015-17) are referred to as second period evaluations in this report.
- **Total evaluation population:** Eligible evaluations for the pre-strategy/early implementation period (i.e. 2009-2013⁴) amounted to 20 evaluations and 28 evaluations for the 2015-2017 period, of which one was disregarded⁵ to avoid any conflict of interest because this evaluation was conducted by Universalialia. Therefore, the total number of eligible evaluations was 47.
- **Dimensions to be covered:** With the aim of conducting as comprehensive an assessment as possible of each evaluation, the quality assessment had two dimensions: i) the first was based on information found in the evaluation summary report; and ii) the second involved review of other associated products for additional information on various aspects of the assessment criteria that were not found in the summary report, e.g. details on design and methodology, complementary information in the annexes, etc. IDEV limits the number of pages of summary reports to a maximum of 20 and the reports' annexes to a maximum of 20 pages as well.

3 Methodology

3.1 Meta-Analysis Design and Data Collection

This quality assessment assignment involved a meta-analysis of various evaluation products and was guided by OECD-DAC quality standards for evaluation⁶ and the Evaluation Cooperation Group's good practice standards for evaluation⁷. It essentially comprised a desk review of the sample of 22 evaluations, including summary reports and corresponding documents such as full technical reports, ToR, approach papers, inception reports, and any relevant documentation available. Quality assessment findings and recommendations were presented to IDEV personnel and further refined during an interactive workshop held in Abidjan on March 27, 2018.

Initial exchanges between IDEV and Universalialia staff assigned to this mandate ensured a common understanding of the assignment. Importantly, it was understood that this assessment was primarily a constructive learning exercise for IDEV and not meant to focus on which evaluations overseen by which task managers scored the highest or the lowest, although information on the highest scoring evaluations

⁴ While evaluations conducted in 2013 were part of the list of eligible evaluations, the assessment team chose to exclude these from the sample on the basis that they coincided with the beginning of the strategy implementation period and therefore did not represent a true "before" portrait for period comparison.

⁵ African Development Bank, *Morocco: Evaluation of the Bank's Country Strategy and Program 2004-2014 – Summary Report*, IDEV Country Strategy Evaluation, September 2016.

⁶OECD Development Assistance Committee, *DAC Guidelines and Reference Series, "Quality Standards for Development Evaluation"*, 2010.

⁷Evaluation Cooperation Group, *Big Book on Evaluation Good Practice Standards*, November 2012; Evaluation Cooperation Group, *Good Practice Standards – Country Strategy and Program Evaluations*, 2008.

was shared with IDEV management to serve as examples of good practice. In this spirit, the analysis of ratings presented in this report does not refer to specific evaluations.

Initial exchanges also provided an opportunity to finalize the sample and fine-tune the methodology and corresponding quality assessment tool which Universalialia then presented in an inception report. Universalialia's overall approach to the assignment was participatory and involved the sharing of views, tasks, and consensus on ratings among the three members of the assessment team.

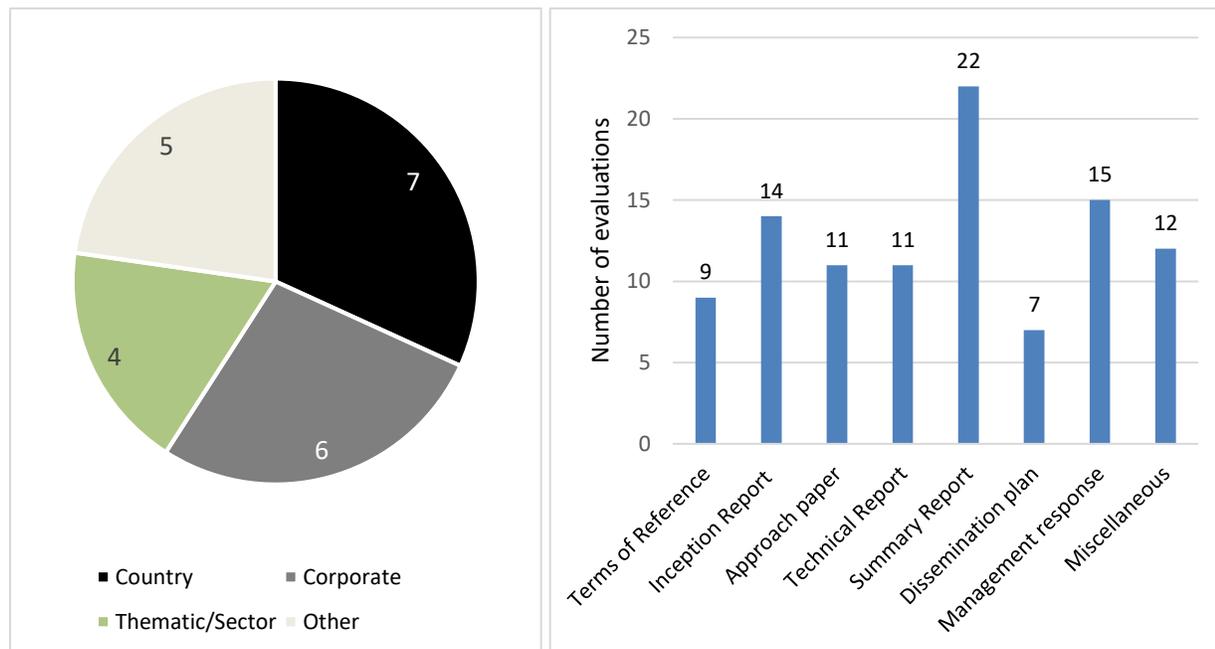
3.2 Sampling, Template, and Analysis Methods

3.2.1 Sampling

In selecting the sample, the assessment team attempted to match types of evaluations between the two periods under review for a representative sample that would allow comparison over time. The team also attempted to achieve proportional representation of the eligible evaluations from each period. The evaluations selected include nine evaluations produced between 2009 and 2011, representing 45% of the eligible evaluations for the early period (2009-2013), and 13 evaluations from the second period, representing 48% of the eligible evaluations for the later period (2015-2017). Other factors that influenced the selection of evaluations representing each period were: prioritize the earliest and latest possible evaluations from each period respectively where change over time would be more likely to be observed; while attempting to match the types of evaluations selected from each period, also include more evaluations in the "others" category from the later period because more evaluations in this category were produced in the later period than in the earlier period; and include evaluations representing both English-speaking and French-speaking countries.

Thus, mindful that evaluations published in 2015 actually began in 2014 when IDEV's evaluation strategy was in its early phase of implementation, the team was careful to select evaluations from that year only when necessary, namely where the portfolio lacked evaluations for subsequent years. Only two evaluations from 2015 were selected. Since most of the 47 eligible evaluations were in the categories "corporate" evaluations (n=11) and "country" evaluations (n=17), six and seven evaluations were selected from each category respectively. Although the category "thematic/sector" evaluations also had a significant number of eligible evaluations overall (n=10), four evaluations were selected from this category because few evaluations were produced in the 2015-2017 period against which to compare them. Finally, five evaluations were selected in the "others" category representing four different types of evaluations (synthesis, regional, cluster, and impact). Three recent evaluations within this category were not comparable with evaluations from the first period but like all other evaluations, were assessed primarily against international benchmarks. Figure 3.1 below shows the distribution of evaluations in the sample by type and includes the number of evaluations for which each evaluation product was found. The "Miscellaneous" category includes various documents such as case studies, orientation/concept notes, progress reports, background papers, explanatory notes, one manual, memoranda, and various project results assessments for country evaluations.

Figure 3.1 Distribution of evaluations by type and evaluation products (n=22)



3.2.2 Assessment Template and Criteria

The quality assessment involved using a standard template developed and agreed upon with the IDEV team during the inception phase. The template includes 26 QA criteria, elements of which are outlined for each in accordance with international evaluation standards, namely the OECD-DAC quality standards and the Evaluation Cooperation Group (ECG) guidelines (see template in Appendix II).⁸

The template is divided into three parts. The first part is essentially descriptive, indicating basic information about the evaluation: title, type, year of delivery, documents reviewed. The second part is divided into four sections representing clusters used to structure the analysis across a series of assessment criteria as follows:

- i) **Cluster 1:** seven criteria related to the context, object, purpose, objectives, and scope of the evaluation;
- ii) **Cluster 2:** four criteria related to the design and methodology of the evaluation;
- iii) **Cluster 3:** five criteria related to the evaluation process; and
- iv) **Cluster 4:** 10 criteria related to clarity and robustness of reporting.

Each criterion is numbered and includes an explanation in italics indicating particular elements assessors should look for when determining their assessment.

⁸ The template was first developed in Word, as presented in Appendix II, then converted to Excel to facilitate data entry and subsequent compilation and analysis.

The criteria incorporate both quantitative and qualitative dimensions, i.e. the assessors determined how many of the elements identified in relation to each criterion were found in the evaluation and assessed elements found for clarity and relevance, coherence with other criteria, etc. Each criterion was rated using a 4-point scale described below and defined in table 3.1. Each section in turn provides for an average assessment rating and balanced appraisal of quality based on ratings of the individual criteria for that cluster. The third part of the template represents an overall assessment of the evaluation to summarize its main strengths and weaknesses and including a percentage of all criteria rated “satisfactory” and “highly satisfactory” across the clusters.

3.2.3 Rating Scale and Rationale for Ratings

Definitions for each rating are provided in Table 3.1 below. In a few instances where a criterion or certain elements of a criterion were not applicable, e.g. for some evaluations in the “others” category, this was noted as not applicable and excluded from the rating. The overall rating for each of the four assessment clusters reflects an average of the corresponding individual criteria ratings using the same 4-point scale and was colour-coded⁹. Because all of the criteria are weighted equally, the overall cluster ratings were calculated as an average of all individual ratings, i.e. all rating points added up and divided by the number of criteria in the section. Scores with multiple digits were rounded up or down to one decimal place and the corresponding colour code reflects the nearest whole number rating. However, overall evaluation assessments in the third part of the template reflect a percentage of individual criteria ratings that were rated “satisfactory” or “highly satisfactory” (3 and 4 scores) only and corresponding comments summarize the main strengths and weaknesses of the evaluation. Percentages were rounded up or down to one decimal place.

Table 3.1 Rating Scale

RATING (CRITERIA)	DEFINITION		RATING (CLUSTER)
4	Highly Satisfactory	The evaluation meets all or nearly all elements of the assessment criterion with no or negligible shortcomings.	3.5 – 4
3	Satisfactory	The evaluation meets the majority of the elements of the assessment criterion with some shortcomings.	2.5 – 3.49
2	Unsatisfactory	The evaluation only meets a few of the elements of the assessment criterion with significant shortcomings.	1.5 – 2.49
1	Highly Unsatisfactory	The evaluation meets almost none of the elements of the assessment criterion with severe shortcomings.	1 – 1.49
—	Not applicable	The criterion or certain elements of the criterion are not applicable	Not rated

Assessor comments accompany all ratings, providing a rationale for the assessment of each criterion and for the overall cluster ratings, and allowing for clear comparison of the strengths and weaknesses of the evaluations between the two periods covered. All team members came to a common understanding of

⁹ 1 = red; 2 = yellow; 3 = light green; 4 = dark green.

the assessment criteria to ensure that all took the same approach. Embedded in the criteria are quantitative and qualitative elements that needed to be weighed to determine a rating for each while taking into consideration the documents, e.g. summary report, technical report, inception report, etc., in which the information was found or not found. An effort was also made to avoid penalizing the same element twice or more. For example, if no evaluation purpose statement was found (criterion 3) and this criterion was consequently rated poorly, other criteria that refer to “coherence with the purpose”, e.g. criteria 4, 5, and 6, were not rated poorly on the basis that coherence with the purpose could not be assessed due to a missing purpose statement. Instead, these criteria were assessed in terms of internal coherence and clarity, e.g. criterion 4, and/or coherence with the evaluation’s objectives, e.g. criteria 5 and 6.

Moreover, because the assessment criteria reflect recognized international standards for evaluations, where information deemed by the assessors to be important for a basic and adequate understanding of the evaluation, and for conformity with international standards, was not found in the summary report but may have been found either in the full technical report, the annexes to the summary report, or just in the executive summary, ratings were downgraded from “highly satisfactory” to “satisfactory” (4 and 3 ratings respectively) on the basis that such information should have been included in the main text of the summary report.¹⁰

The following table provides examples of ratings from 1 to 4 for a selected criterion to illustrate how ratings were justified in a few cases. It should be noted that these examples drawn from the QAs are simply an illustration and are not meant to fully represent the kind of information that is necessarily expected to be found for this criterion and, in any case, some comments may not be relevant depending on the type of evaluation. For example, some comments do not apply to corporate evaluations. As mentioned at the beginning of this section, where a criterion or certain elements of a criterion were deemed not applicable, this did not affect the rating.

Table 3.2 Case examples of four different ratings and assessments for one criterion

CRITERIA	RATING/ASSESSMENT			
	1	2	3	4
<p><u>Cluster 1</u></p> <p>1. The background/context of the object of the evaluation is clearly described and includes all relevant factors.</p> <p><i>Key social, cultural, political, economic, and institutional factors that are relevant to the object of the evaluation.</i></p>	<p>The evaluation does not include a section which would contextualize the Bank's support over the period under review. It would be important to include a section on context, perhaps presenting general information on key social, cultural, political and economic trends in Africa</p>	<p>While the inception report includes some relevant context information that provides a rationale for the evaluation, the summary report provides very little information about the context, with information strictly focused on the institutional</p>	<p>The summary report includes a section on country context, which provides a detailed overview of the geographical and political context, as well as an overview of national development strategies, the country's economic growth and the institutional factors hindering economic</p>	<p>The evaluation provides relevant information on the context in AfDB, in the continent in general, and in specific countries concerned and how the situation has evolved during the period covered by the evaluation. The evaluation discusses the context in which initiatives were implemented,</p>

¹⁰ There are only a few instances of this in the quality assessment.

CRITERIA	RATING/ASSESSMENT			
	1	2	3	4
	during this period. This would help the reader better understand in which context the Bank's assistance was provided.	context. The technical report background section introduces the evaluation but does not include a description of the context. Some basic contextual elements described in the inception report should be included in the summary report.	development. Included is a brief description of successive national programs that informed Bank strategies and programs during the period under review. Information on the social and cultural context in the country, including the situation of ethnic and marginalized groups, could be provided in the main text of the summary report since this is relevant information in the context of this evaluation, but key socio-economic indicators are included in the annexes.	including recent changes in the international and regional context, developments in the international aid architecture and the international economic environment. This section also discusses political factors including conflict and instability (e.g. Sierra Leone, DRC) or political crisis (e.g. Cote d'Ivoire, Zimbabwe). The section also discusses how the increased focus on aid effectiveness (leading to the Paris Declaration) affected the object of the evaluation. Finally, the section discusses the Bank's institutional framework (structures and policies), in which the object has evolved.

The QA process involved three steps, as illustrated in Figure 3.2 below. Firstly, each evaluation was assessed in parallel by two assessors. Secondly, the two assessors of each evaluation discussed their ratings/comments and consolidated them into a single assessment for each evaluation. Thirdly, and once all evaluations had been consolidated, the full team reviewed all assessments to ensure overall consistency and quality, making minor changes to some ratings of consolidated reviews and resulting in harmonized assessments across all criteria.

Figure 3.2 Three-step quality assessment process

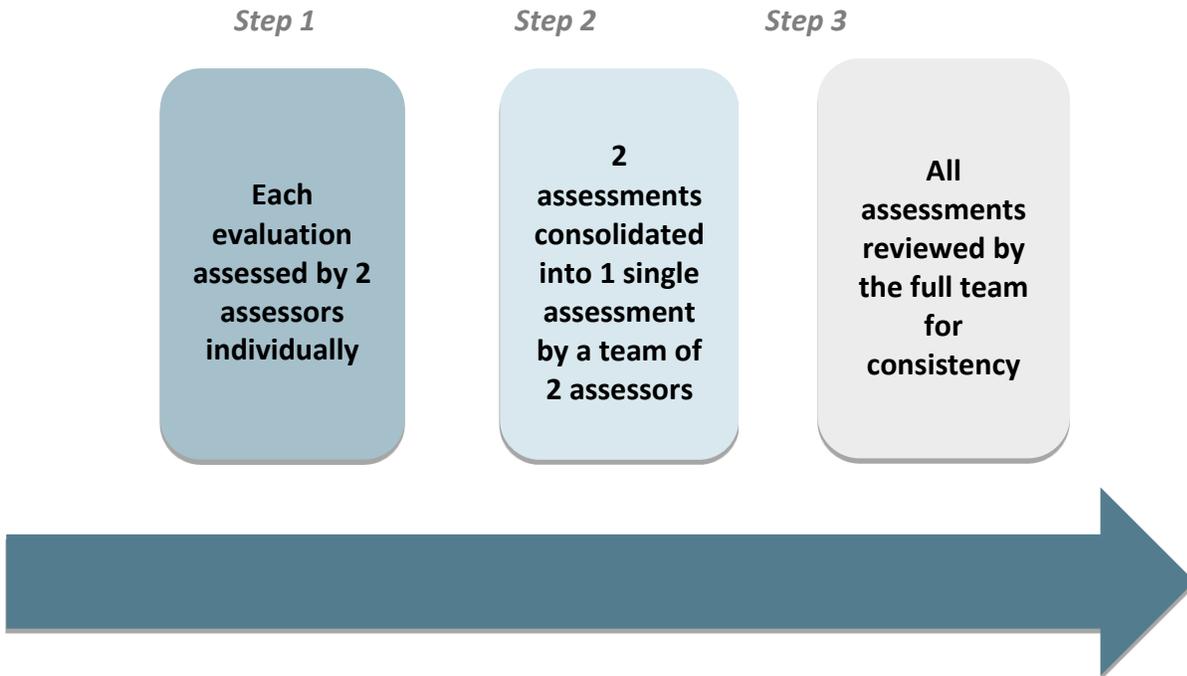
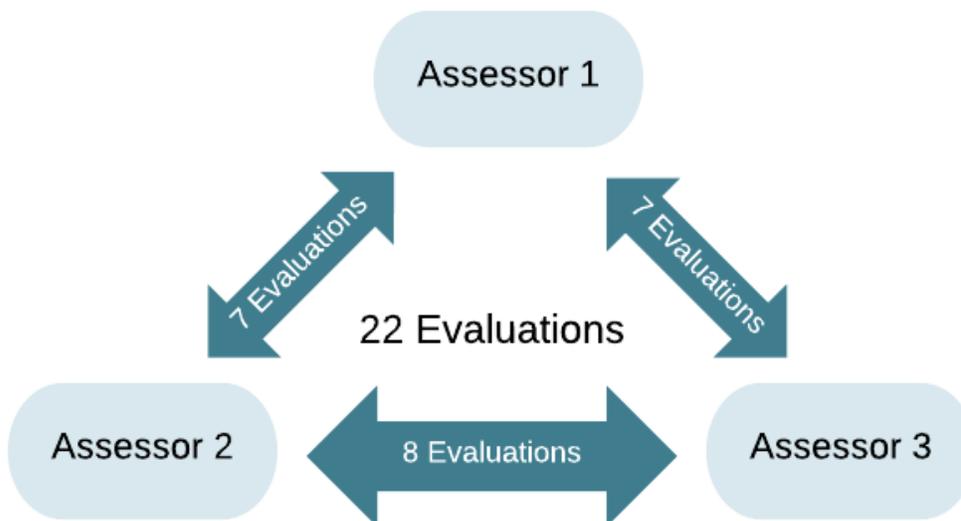


Figure 3.3 below illustrates how the evaluations were distributed among the assessors. Assessors 1 and 2 individually assessed then consolidated their assessments of seven evaluations while Assessors 1 and 3 individually assessed then consolidated their assessments of another seven evaluations and Assessors 2 and 3 individually assessed then consolidated their assessments of the remaining eight evaluations for a total of 22 evaluations covered. Thus, one assessor reviewed a total of 14 evaluations while two assessors reviewed a total of 15 evaluations. All three assessors then discussed the ratings of all 22 evaluations to ensure overall consistency.

Figure 3.3 Distribution of evaluations for among assessors



3.2.4 Quantitative and Qualitative Analysis

Upon completion of the third step, the review team employed mixed data analysis methods to inform the findings of the quality assessment, as follows:

- **Qualitative descriptive analysis** to summarize the key characteristics of the evaluations (e.g., type and year of evaluation, type of documents available and reviewed for each evaluation);
- **Quantitative descriptive analysis** to determine the number of evaluations associated with the key characteristics identified;
- **Comparative qualitative and quantitative analysis** to compare findings (ratings and assessments) for each of the 26 QA criteria across all evaluations and specifically between first period evaluations and second period evaluations within and between different types of evaluations.
- **Cluster analysis** to compare assessments for criteria within and across the four groups of criteria (clusters) for all evaluations and specifically between first period and second period evaluations and between different types of evaluations;
- **Content analysis** was used to identify common trends, themes, and patterns that emerged for each of the key units of analysis or, conversely, to highlight opposite trends.

In sum, the analysis aimed to identify the strengths and weaknesses of the evaluations, emerging themes, and observed changes over time.

3.3 Limitations and Mitigating Actions

- **Limited data available for some evaluations.** The desk review was necessarily constrained by the amount of information available which varied from one evaluation to another. While IDEV provided all the documents it had for each evaluation in the portfolio, the full suite of documents was available for some while others, in particular evaluations from the first period, had fewer documents available. It is possible that some documents for the first period evaluations were lost or that these evaluations produced fewer documents. Since all evaluations had at least the summary report, this limitation was mitigated by focusing the quality assessment primarily on this report. Only where certain criteria necessarily involved comparison of information in summary reports with other products or where there were gaps in information in summary reports preventing an adequate rating of some assessment criteria did assessors look into other documentation available to fill the gaps.
- **The “static” nature of the desk review that may not account for current developments.** New policies or orientations in programming and evaluation may be in place or in development at IDEV-AfDB since the last evaluation in the sample was produced in 2017 and are therefore outside the scope of this assessment. This assessment provides a retrospective look at the differences in quality of evaluations between two specific periods of time. While it may not account for changes in evaluative practice that may have been introduced recently, it nonetheless contains information that can inform current and future initiatives. Moreover, where changes or clarifications in evaluative practice were brought to the attention of the assessment team representative at the learning workshop in Abidjan, these have been noted in the report.

- **Limited statistical significance.** While the sample is representative of the various types of evaluations conducted by IDEV, due to the small sample size, the results of this assessment were not intended to be statistically significant.
- **Inherent difficulties in reconciling quantitative and qualitative elements within one rating.** The assessment criteria embedded both quantitative and qualitative elements that had to be balanced in order to determine a fair rating for each criterion. This was facilitated somewhat by identifying in the template various elements that assessors should consider for each criterion both quantitatively (number of elements) and qualitatively (clarity, relevance, coherence, etc.). These elements were later supplemented by more specific rating cues to assessors for each criterion.
- **Challenges posed by the atypical nature of some evaluations,** namely those in the “Others” category like cluster or synthesis evaluations, made it more difficult to rate the quality assessment criteria that were developed based on standard evaluation criteria more suited to formative or summative evaluations of specific interventions. Where certain criteria did not clearly apply to these evaluations, a non-applicable “rating” was given that did not negatively affect average cluster ratings.
- **Limited comparability between the first and second periods for some types of evaluations within the “others” category** because these types of evaluations were not produced in earlier years. This was not a significant limitation since the quality of all evaluation products was assessed against the same criteria in conformity with international benchmarks.
- **Potential assessor bias** in the interpretation of both the rating criteria and the information found in evaluation documents was avoided through initial team consultations to ensure a common understanding of the QA criteria, followed by the rigorous three-step process of data analysis that involved two assessors individually reviewing each evaluation, then coming together to consolidate their ratings, followed by the full team of assessors meeting to harmonize ratings for each criterion in all 22 quality assessments. Moreover, after a few assessments had been completed by each two-person team, rating cues for assessors were added next to each criterion in the template to address particular cases.
- **Labels and definitions in relation to the 4-point rating scale did not allow for ratings between “unsatisfactory” and “satisfactory”** and could therefore not accommodate a more nuanced assessment of some criteria, deemed in some instances to be somewhere in between. To mitigate this, assessors relied on their judgment to decide whether the information merited one or the other of these ratings, i.e. leaned more toward an unsatisfactory or a satisfactory assessment.

4 Main Findings

4.1 Overall Analysis of Quality of Evaluations

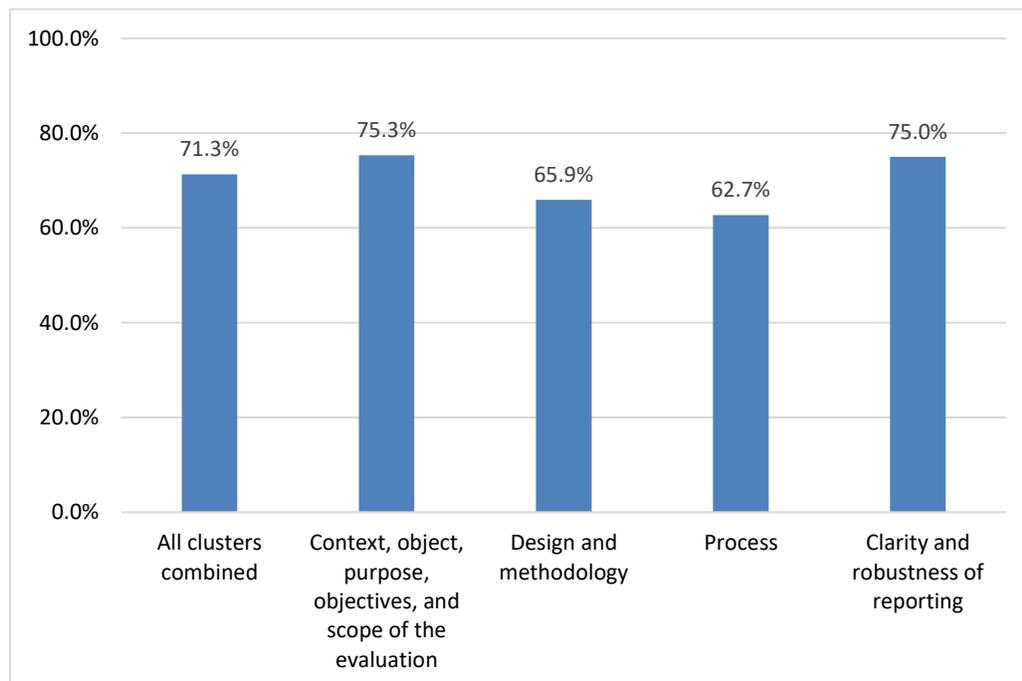
4.1.1 Overall Ratings by Cluster

Since no average overall rating was calculated for each evaluation, the overall assessment of the quality of evaluations is reflected in the percentage of “satisfactory” and “highly satisfactory” ratings for each evaluation as well as the average rating determined for each cluster of criteria, as defined in the methodology section of this report.

Finding 1: For all evaluations combined, the majority of quality assessment criteria were rated “satisfactory” and “highly satisfactory”.

Slightly over 70% of QA criteria received a rating of “satisfactory” or “highly satisfactory” with a fairly even distribution of these ratings among the four clusters of criteria showing only a 12.6% difference between the lowest and highest figures (see Figure 4.1). Criteria related to the context, object, purpose, objectives, and scope of the evaluation scored best at 75.3% in comparison to other clusters, very closely followed by criteria related to clarity and robustness of reporting, 75% of which were rated “satisfactory” and “highly satisfactory”. Process-related criteria on average received fewer ratings at the upper end of the scale.

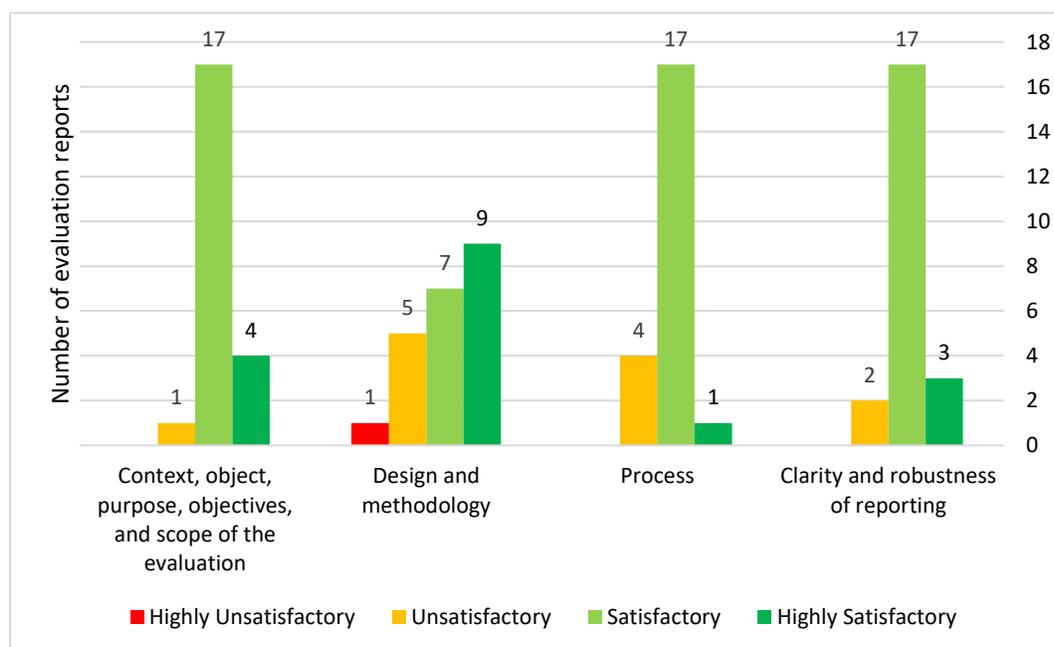
Figure 4.1 Percentage of criteria rated 3 and 4 by cluster, all evaluations



Finding 2: The most frequent rating for all four clusters of criteria combined was “satisfactory” with design and methodology showing the greatest variation in ratings.

The most frequent rating reported was “satisfactory”, reflecting an overall positive picture of the quality of evaluations assessed (see Figure 4.2). Clusters 1 and 4 received the highest ratings overall, with 17 evaluations rated satisfactory on average for these clusters, while cluster 2 shows the most variation in average ratings.

Figure 4.2 Average cluster ratings, all evaluations (n=22)



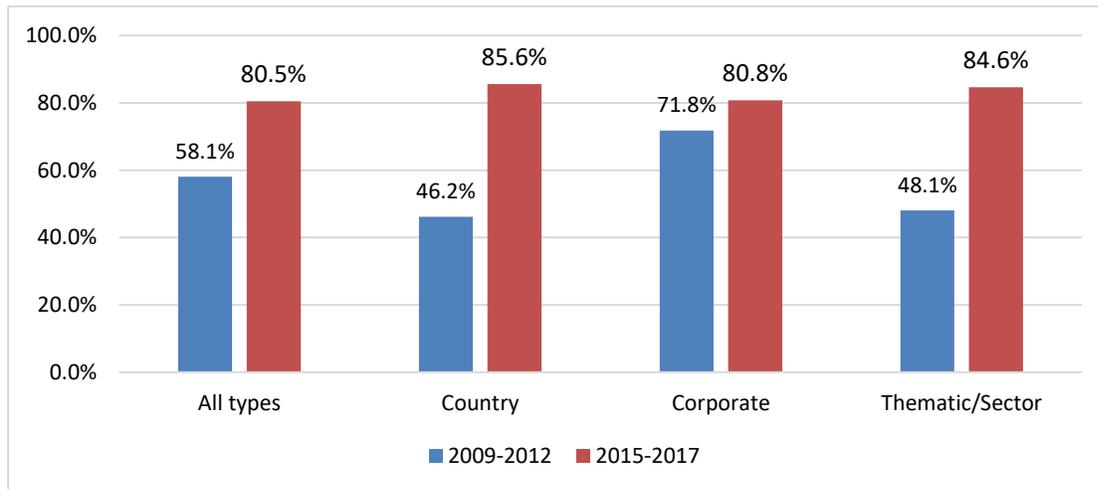
4.1.2 Overall Comparison across Types of Evaluations between First and Second Periods

This section compares criteria and average cluster ratings between the first and second periods and across types of evaluations.

Finding 3: There is a marked improvement in the quality of evaluations over time from 2009-12 to 2015-17.

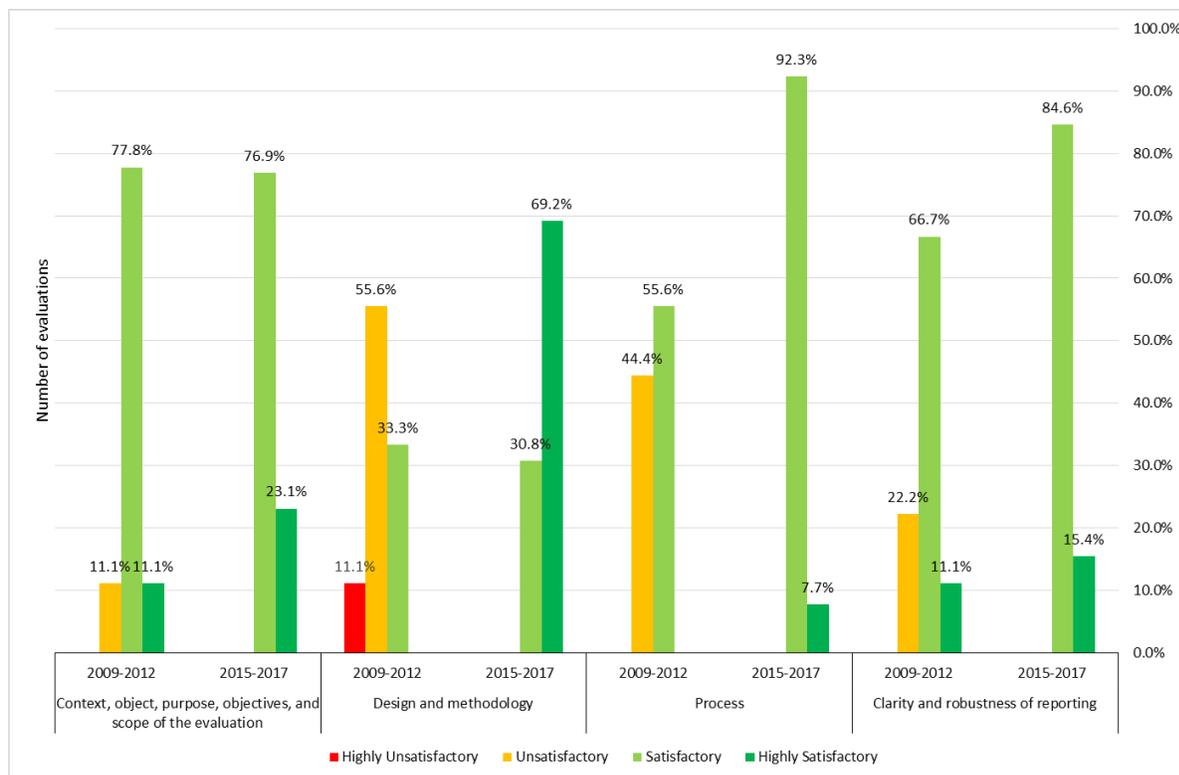
Comparing assessments of evaluations between the first period and second period, there is a clear improvement in quality over time with country and thematic/sector evaluations showing a sharper increase in quality than corporate evaluations simply because the latter were rated higher than the other types of evaluations for the first period (see Figure 4.3). It should be noted that the one synthesis evaluation in the first period could not be compared by period to the four other evaluations reviewed in the “Others” category.

Figure 4.3 Comparison of percentage of criteria rated 3 and 4 by evaluation period and type



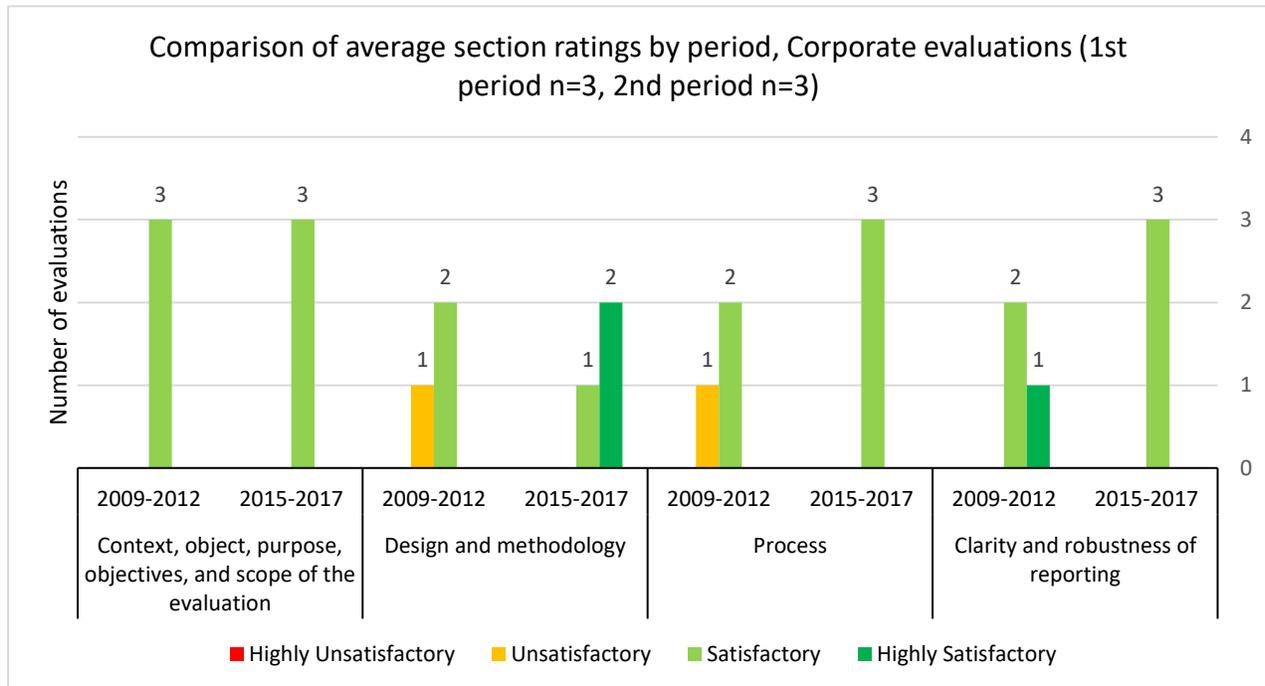
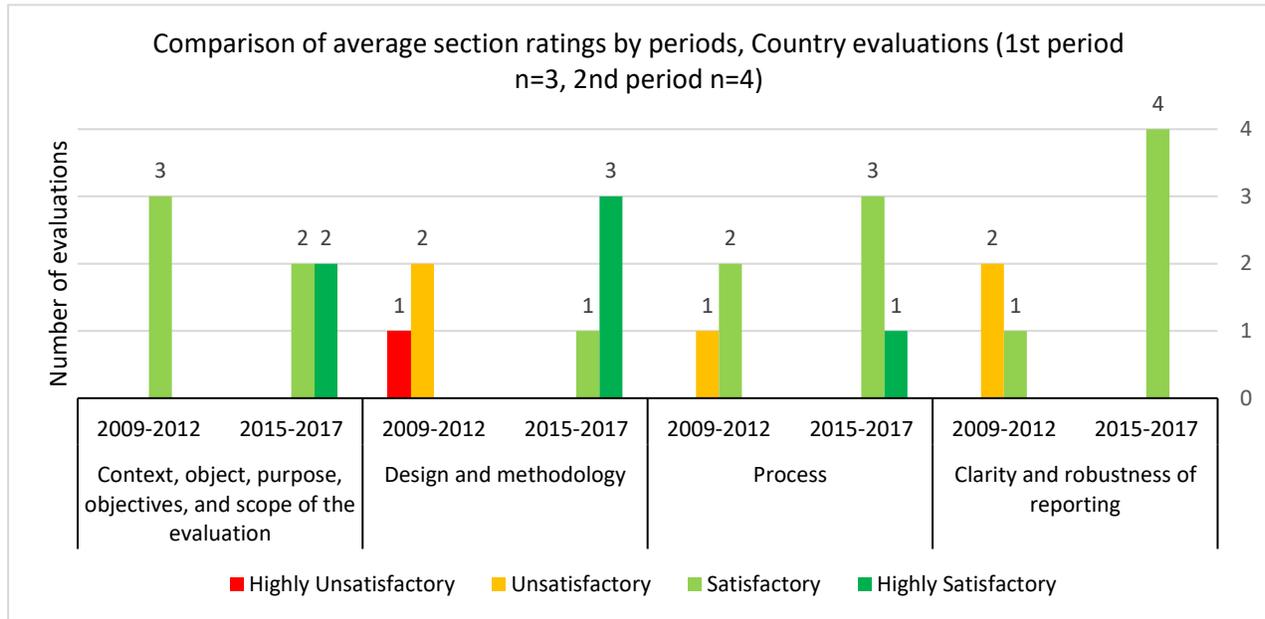
While between 11.1% and 55.6% of first period evaluations received an average “unsatisfactory” rating for each cluster of criteria, all 13 second period evaluations received either a “satisfactory” or “highly satisfactory” average rating for all clusters (see Figure 4.4). The percentage of “satisfactory” and “highly satisfactory” ratings for each of the 26 criteria in all evaluations is presented in Appendix III.

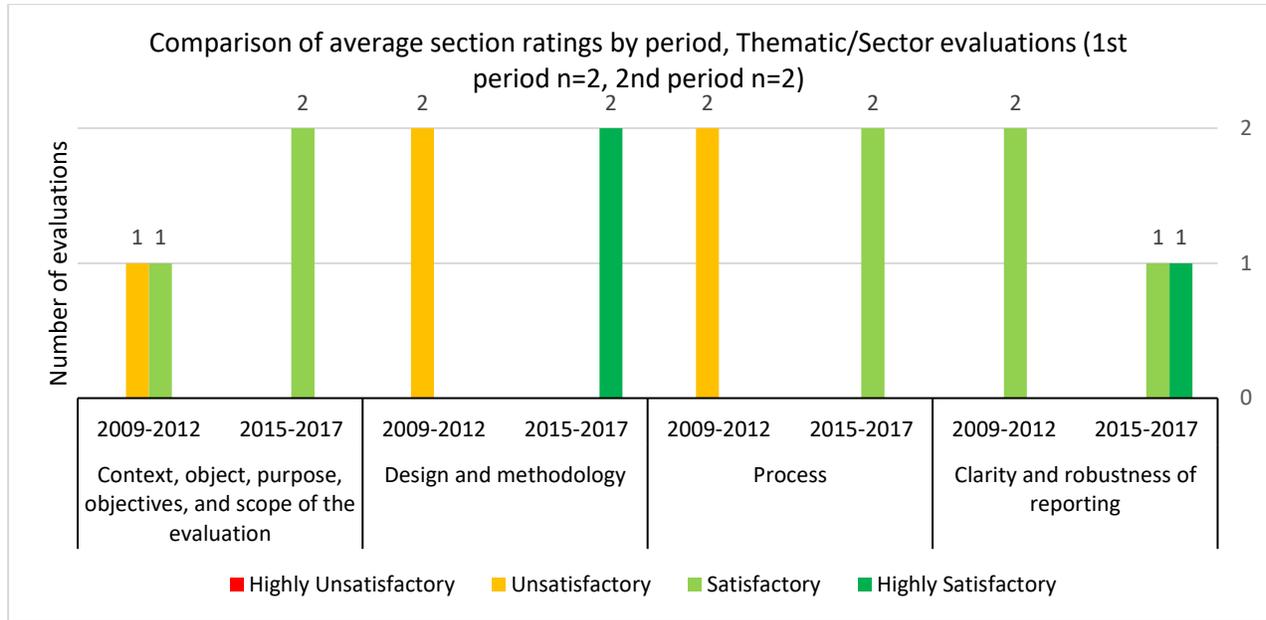
Figure 4.4 Comparison of average cluster ratings between first period and second period, all evaluations (first period n=9, second period n=13)



With the different types of evaluations, except those in the “Others” category, almost equally represented in number between the first and second periods, there was a notable improvement across all types from the first to the second period (see Figure 4.5).

Figure 4.5 Comparison average cluster ratings by type of evaluation, first period and second period





4.2 Comparative Analysis of Strengths and Weaknesses of Evaluations by Criteria between First and Second Periods

This section compares the strengths and weaknesses of the evaluations reviewed for each of the two periods, reflecting on emerging themes through the more specific assessment of individual criteria in each of the four clusters.

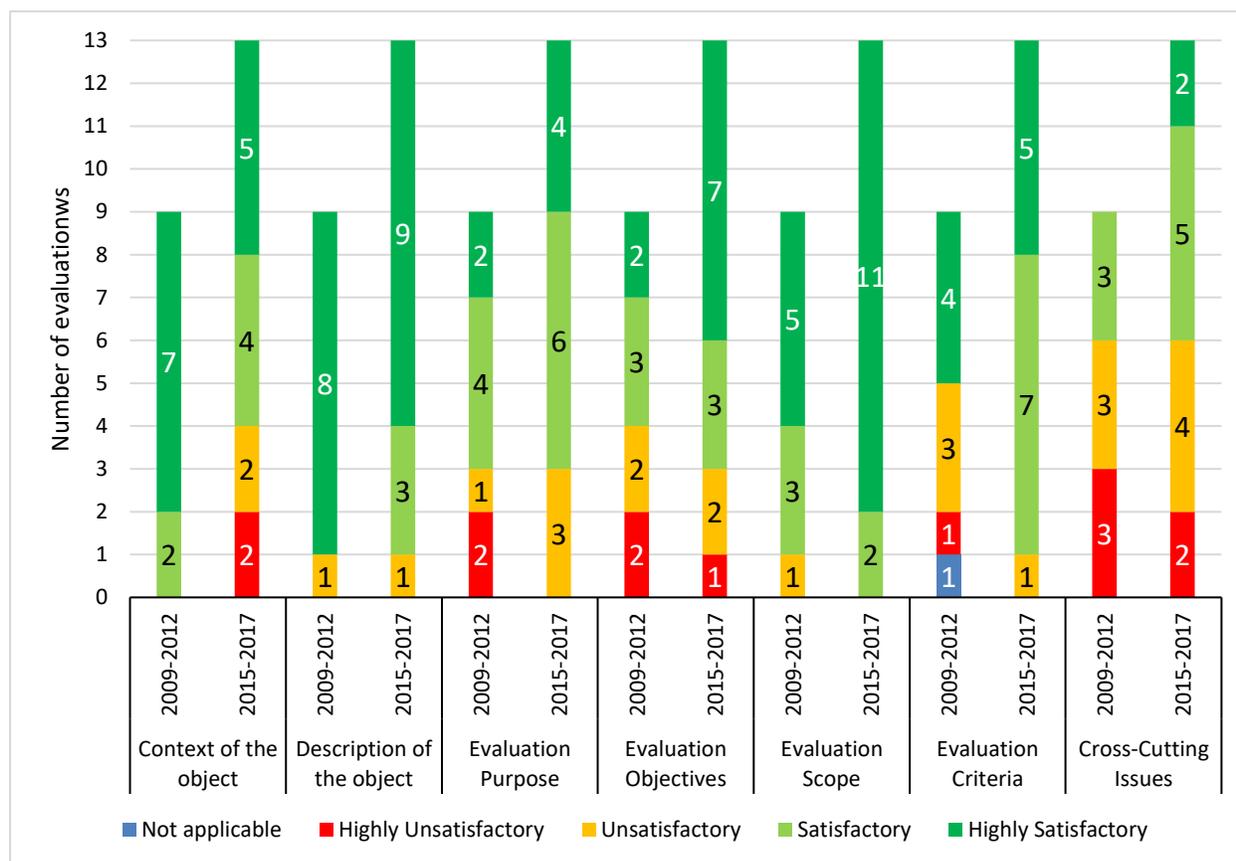
4.2.1 Context, Object, Purpose, Objectives, and Scope of Evaluations

Finding 4: Description of the object of the evaluation and scope of the evaluation are the highest rated criteria, and cross-cutting issues the lowest rated criterion, within this cluster in both periods under review.

Among the seven criteria related to the context, object, purpose, objectives, and scope of the evaluation, the description of the object and of the scope of the evaluation were the best rated on average (see Figure 4.6). Comparing ratings for individual criteria between first and second period evaluations, while first period evaluations show slightly stronger ratings in relation to the description of the context of the object of the evaluation, second period evaluations show a notable improvement in ratings related to the clear identification of evaluation criteria coherent with the purpose of the evaluation.

Most evaluations reported on the timeframe of the object of the evaluation, its main objectives, strategies, and budget. A good practice identified relates to how certain evaluations included a thorough description of the object of the evaluation and its context. For example, some evaluations provided details on similar assessments carried out before and how the Bank had responded to these assessments.

Figure 4.6 Cluster 1 Ratings: Comparison between Periods (first period n=9, second period n=13)



Coverage of cross-cutting issues in this cluster received the lowest rating on average, with almost 50% (6/13) of second period evaluations rated either “unsatisfactory” or “highly unsatisfactory”, and only two out of 13 second period evaluations (15%) rated “highly satisfactory”, on this criterion. The main reason for rating this criterion “unsatisfactory” or below was the fact that cross-cutting issues were not addressed in the description of the object of the evaluation and its context, nor in the scope of the evaluation. For example, with respect to gender equality (GE), some evaluations only partially addressed this issue by including sex-disaggregated indicators in the evaluation matrix, but did not include a narrative explanation of how the GE dimension would be addressed in the evaluation process. Good examples include evaluations that provided contextual information on cross-cutting issues in the description of the background and object of the evaluation, considered cross-cutting issues in the evaluation questions, and included a statement indicating an explicit intention to address cross-cutting issues in the evaluation process.

The majority of evaluations were rated “satisfactory” and “highly satisfactory” in relation to the description of the evaluation purpose and objectives with notable improvement again observed with respect to these criteria over time from the first to the second period. However, it should be mentioned that several evaluations, even in the second period, did not include in the statement of purpose the relevance of their timing, the intended users or use. Moreover, some evaluations did not clearly distinguish between evaluation purpose and objectives.

In terms of variance by type of evaluation, country evaluations are slightly better rated than other types of evaluation, being the only type of evaluation having “highly satisfactory” ratings for this cluster. Thematic/sector evaluations are slightly less well rated, as they include one evaluation rated “unsatisfactory” with respect to this cluster (see Figure 4.5 in section 4.1.2).

4.2.2 Design and Methodology

Finding 5: Within this cluster, data collection, analysis and sampling is the highest rated criterion, with no evaluations showing a “highly unsatisfactory” rating on this criterion, and alignment of evaluation questions with criteria and indicators is the lowest rated criterion, although there was significant improvement on this criterion in the second period evaluations.

With respect to design and methodology criteria, the most “satisfactory” and “highly satisfactory” ratings combined were found to be in relation to data collection, analysis, and sampling (see Figure 4.7). However, almost half of the country evaluations, all representing the first period, were weak in their descriptions of methodologies overall (see Figure 4.5, section 4.1.2). Issues concerned poor identification of the evaluation design, the absence of an evaluation matrix or of a theory of change (ToC), weak description of the sampling methods or of the rating scale used to assess performance.

The most notable improvements over time among this cluster’s criteria concern the alignment of evaluation questions with evaluation criteria and indicators in an evaluation matrix and identification of evaluation risks and/or limitations (see Figure 4.7). All evaluations in the first period and 23% (3/13) of the evaluations representing the second period did not satisfactorily present an evaluation matrix. In some recent cases where an evaluation matrix was included, some key elements were missing in the matrix, notably matching questions with appropriate indicators or with the data collection methods and sources. This concerned evaluations in the “Others” category in particular. While all second period evaluations rated satisfactorily on data collection, sampling, and analysis as well as the identification of evaluation limitations and mitigation strategies, some 30% (4/13) did not satisfactorily identify the evaluation design.

Case Example: The African Development Bank’s Human Resource Management Policy and Strategic Directions: A Formative Evaluation

The formative and theory-based evaluation design is clearly described and appropriately aligned with the purpose. A detailed mixed data collection and analysis strategy is described with a comprehensive sample of data sources. Relevant evaluation questions are outlined in a detailed evaluation matrix in annex. Methodological limitations and mitigation strategies are appropriately described.

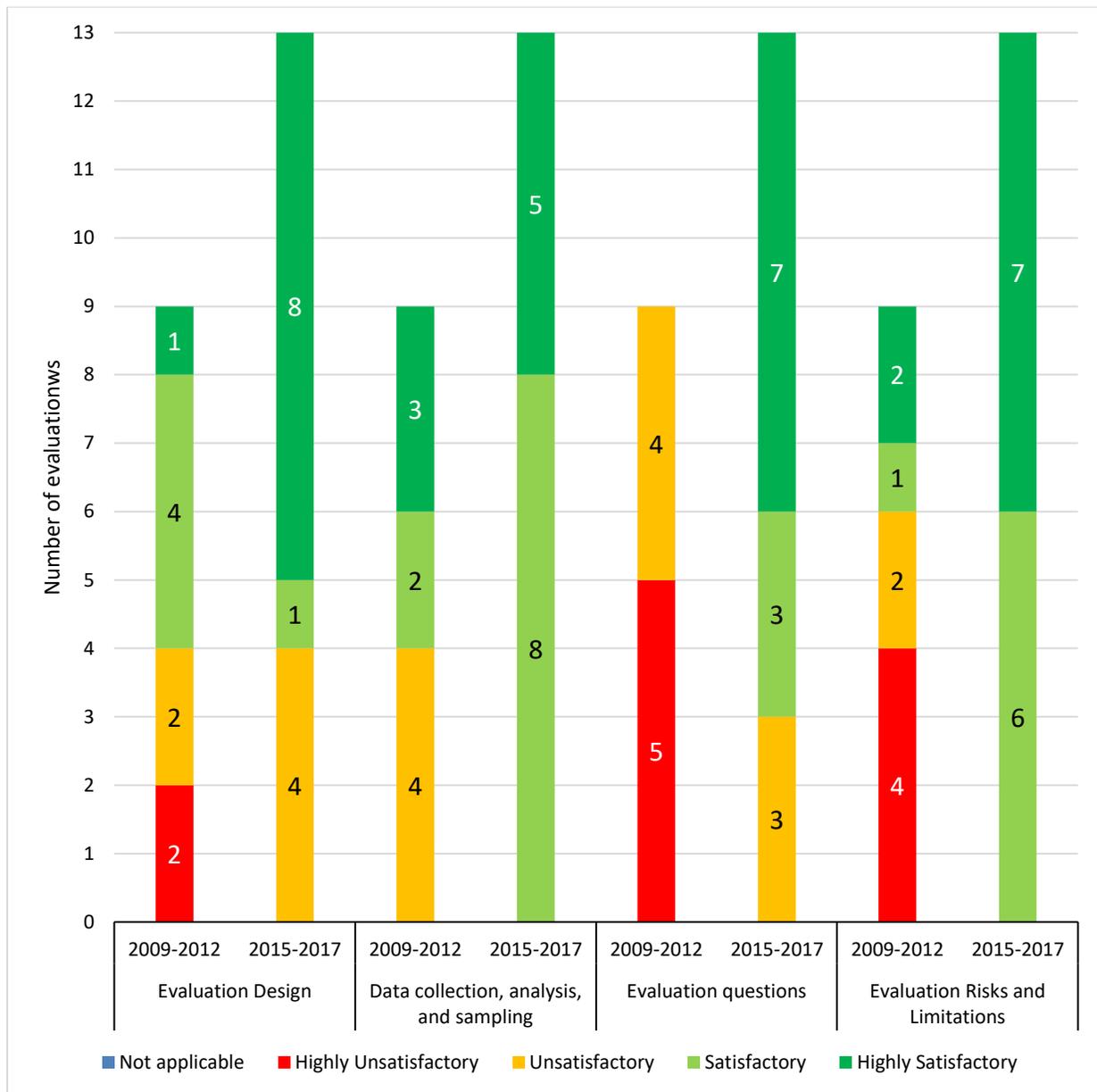
The rating scale used by IDEV evaluations was also found to be problematic. Referring to the examples of one country evaluation and one corporate evaluation, a significant amount of effort appears to have been put into explaining how ratings for each level of the 6-point rating scale used were defined (e.g. CEDR evaluation, Annexes D, G, and H; Ghana evaluation, Annexes C and E).¹¹ Without these definitions

¹¹ Six-point rating scale labeled as: highly unsatisfactory (HU), unsatisfactory (U), moderately unsatisfactory (MU), moderately satisfactory (MS), satisfactory (S), highly satisfactory (HS).

however, ratings would be less clear since one could technically interpret middle-level rating labels (MU and MS) to be one and the same thing. That said, even where these ratings were defined, they sometimes still lacked clarity by failing to clearly set upper limits for the ranges.¹² Such finer-level definitions that specify a quantitative range may be more pertinent for evaluating predominantly quantitative factors or criteria, such as efficiency. It should be noted that IDEV has recently replaced the 6-point rating scale by a 4-point rating scale labeled as follows: “highly unsatisfactory”, “unsatisfactory”, “satisfactory”, “highly satisfactory”. However, while less cumbersome than the previous 6-point scale, this scale is also problematic because it does not allow for a more nuanced assessment between “unsatisfactory” and “satisfactory”.

¹² For example, the definitions for the MU and MS ratings in relation to assessing the relevance criterion in Annex D of the CEDR synthesis evaluation.

Figure 4.7 Cluster 2 Ratings: Comparison between Periods (first period n=9, second period n=13)



4.2.3 Process

Finding 6: The links between evaluation products for each evaluation were logical overall while evaluation principles – independence, impartiality, ethics, coordination and alignment, partnership and capacity development – were not satisfactorily addressed across both periods.

While the quality ratings related to process were quite variable (see Figure 4.8), linkages between evaluation products were found to be strong overall, with 89% (8/9) and 85% (11/13) of evaluations representing the first and second periods respectively rated “satisfactory” or higher on this criterion.

Some repetition of information was noted among approach papers, inception reports, and sometimes ToR for the same evaluations, but IDEV personnel explained that, while not all of these products are necessary for all evaluations, they each serve a distinct purpose depending on the context of the evaluation and the extent to which external consultants are involved. However, IDEV does not have specific guidance on what is minimally expected to be found in each type of document.

Some evaluations were highly rated on consideration of evaluation principles, particularly knowledge sharing through the dissemination of findings. However, while IDEV's policy emphasizes the guiding principles of independence, credibility, impartiality, transparency, and partnership¹³, these principles, in addition to coordination and alignment, and capacity development, were not well, or explicitly, addressed overall in the evaluations reviewed. These principles may have been applied in practice but there is no clear evidence of this. Evaluations that received ratings of "satisfactory" or higher for integrating other OECD-DAC and IDEV evaluation principles improved by 15% (2/13) in the second period from nil in the first period, yet most evaluations still rated "unsatisfactory" or lower in this area. Very few evaluations refer to such evaluation principles and, when they do, the reference is only generic and it is unclear how principles were applied concretely.

Regarding quality assurance, 78% (7/9) of evaluations conducted during the first period rated unsatisfactorily on this criterion simply because quality assurance mechanisms were not addressed in the set of documents reviewed. Conversely, 92% (12/13) of evaluations conducted in the second period rated "satisfactory" or higher on this criterion. All but one evaluation described the quality assurance process. Evaluations that received a "highly satisfactory" rating had internal and external quality assurance mechanisms, such as peer review committees or independent experts which reviewed and commented on evaluation deliverables.

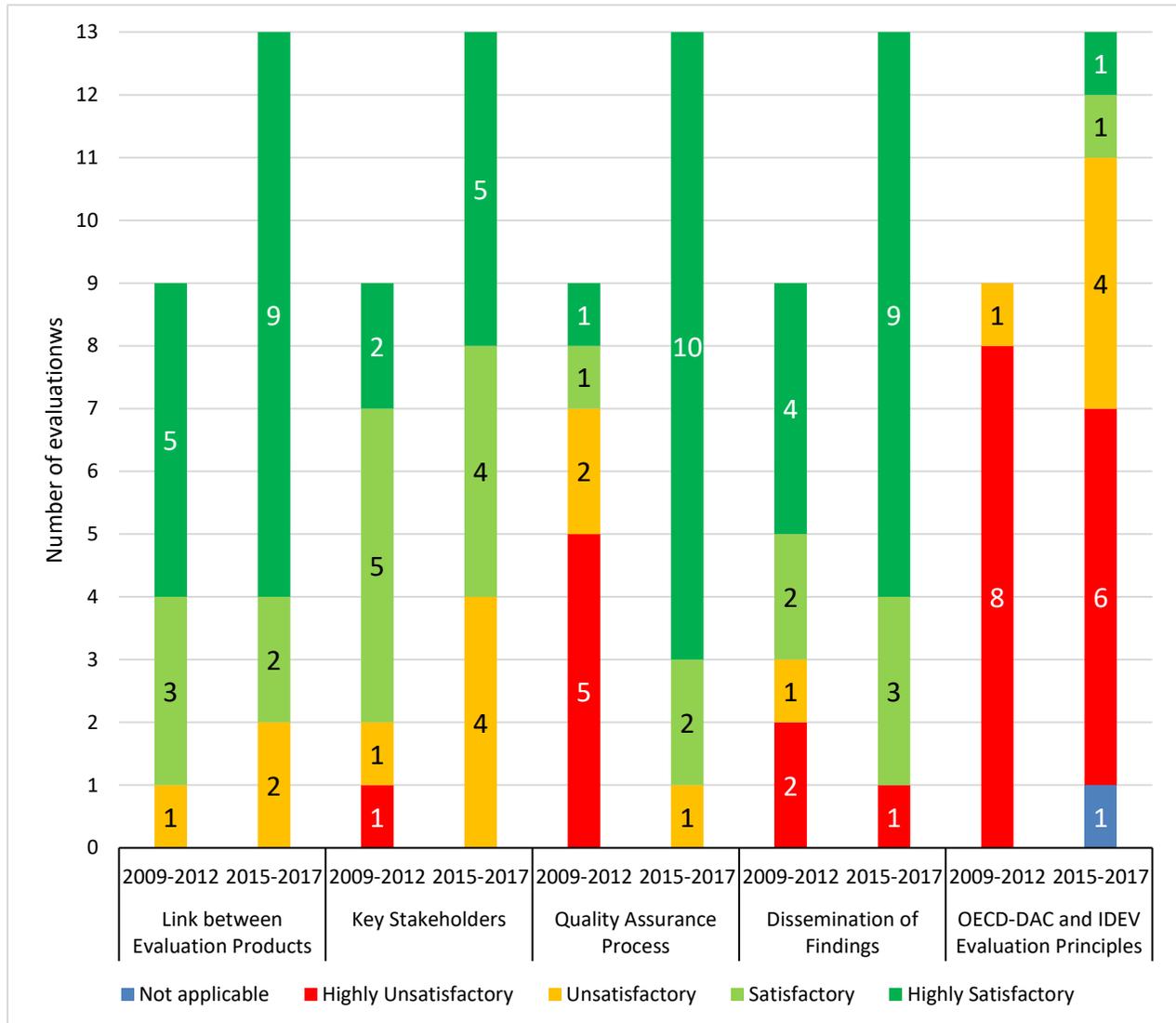
An opposite trend was noted for key stakeholder engagement between the first and second periods, with 78% (7/9) of first period evaluations and 69% (9/13) of second period evaluations rated "satisfactory" or higher on this criterion. Evaluations that scored "highly satisfactory" on this criterion described the process for engaging stakeholders in the evaluation design, as well as in data collection and the validation of findings. Given the importance of stakeholder participation in evaluation, it is unclear why the more recent evaluations did not always systematically identify stakeholders and their roles and contributions in the process.

Case Example: Quality assurance

The *Comprehensive Evaluation of the Development Results of the African Development Bank group 2004-2013*, published in 2016, was identified as good practice because it described internal and external quality assurance. Internally, context factor review (CFR) and project results assessments were reviewed by two staff and coordinated at the central level to ensure consistency. Externally, a panel of senior independent advisors was called upon to provide methodological advice during the design phase, as well as feedback on the draft and final report. This panel also provided a letter, included in the evaluation report, which attest to the quality of the evaluation, confirming that the evaluators have addressed feedback from the panel.

¹³ Independent Evaluation Policy The African Development Bank Group, July 2016, pp. 2-4.

Figure 4.8 Cluster 3 Ratings: Comparison between Periods (first period n=9, second period n=13)



Finally, regarding the dissemination of findings, 92% (12/13) of evaluations conducted during the second period scored “satisfactory” or higher, compared with 67% (6/9) for the previous period. In recent evaluations, the inclusion of a dissemination plan identifying target audiences for various communications products such as briefs, summary reports or PowerPoint presentations is considered good practice. Some dissemination plans also include a timeline and person responsible for conducting evaluation activities.

4.2.4 Clarity and Robustness of Reporting

Finding 7: The large majority of evaluation reports were well structured and included a concise executive summary, presented evidence-based findings, addressed the evaluation criteria and questions, and included relevant recommendations. However, the sources from which data were derived were often not explicitly identified in evaluations from both periods.

Approximately 75% of first period evaluation reports and all second period evaluation reports were rated satisfactorily on criteria pertaining to the structure of the report, the executive summary, the presentation of the findings based on evidence that addressed the evaluation criteria and questions, and the clarity of the recommendations (see Figure 4.9).

However, in the findings reported, across all types of evaluations and all periods, while extensive documentary sources were consulted and some key stakeholders consulted in the evaluation process, with respect to primary sources, the specific stakeholder groups from which evidence was derived were often not identified such that it was frequently unclear what their common or distinct views were and to what extent data collection methods and sources were triangulated.

Most reports systematically addressed the evaluation criteria and questions and included relevant recommendations. However, recommendations usually did not specify a clear timeline for implementation that would indicate the level of priority of each recommendation.

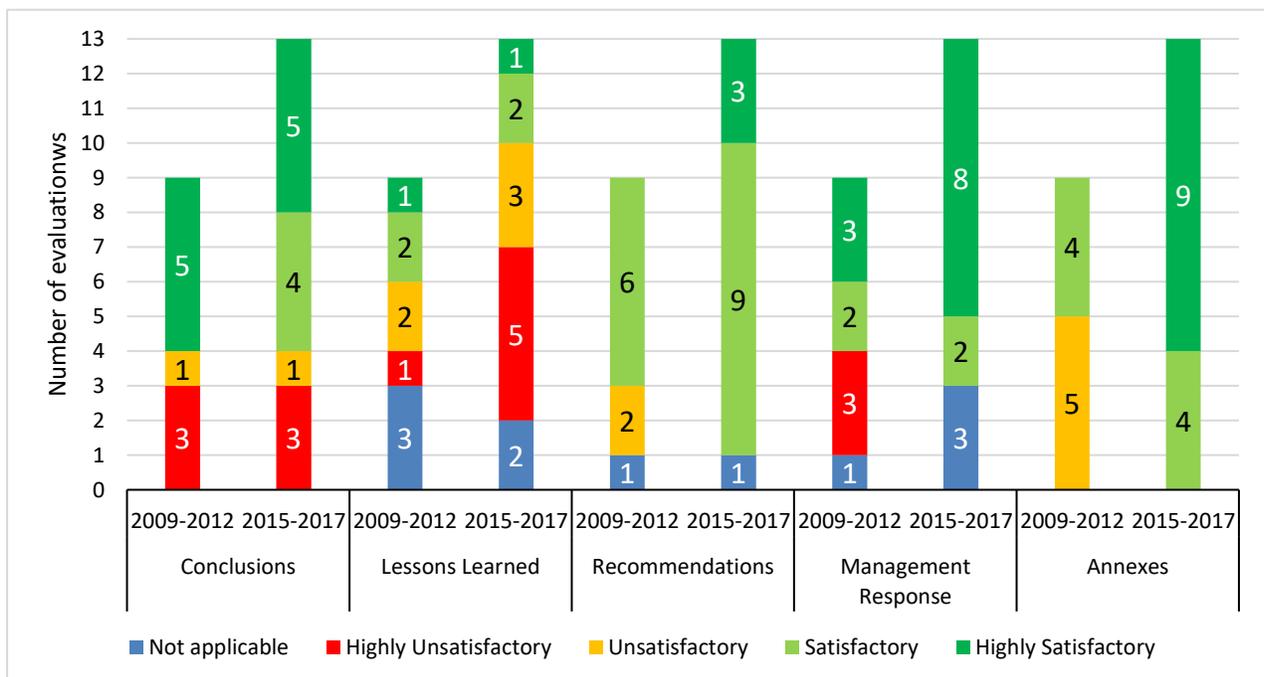
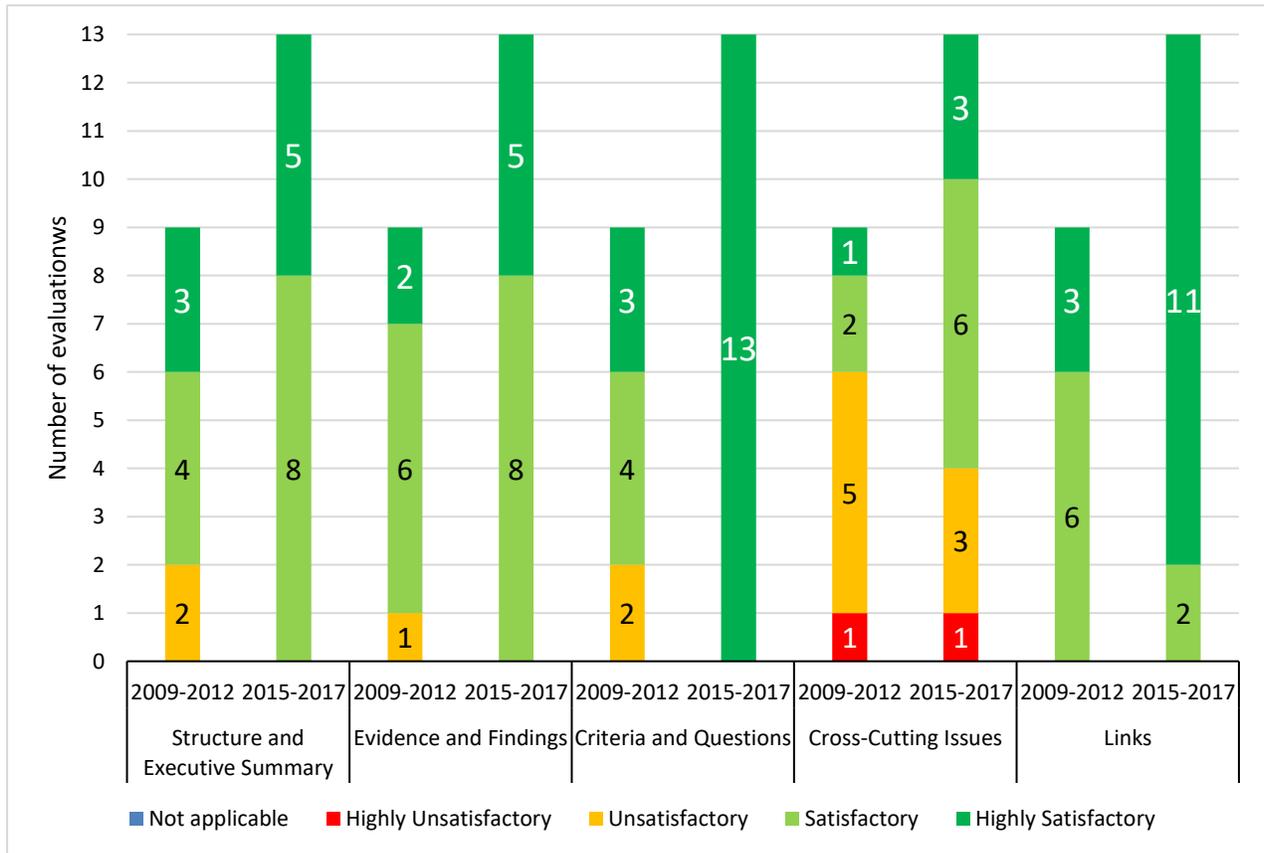
Almost all evaluations from the second period include a management response¹⁴ and all management responses reviewed indicate agreement overall with the recommendations made.

Case Example: South Africa - Country Strategy and Program Evaluation 2004-2015

The evidence is based on multiple lines of evidence, including a portfolio review and consultations with key stakeholders. In the synthesis report, data sources are consistently presented across the report, and the views of different stakeholders allow for a nuanced picture of the Bank's performance. It is clear how the information was triangulated. Data gaps are also discussed. For instance, the synthesis report confirms that there was not enough information to demonstrate whether the Bank managed itself cost-efficiently.

¹⁴ While there may exist a management response for all the evaluations reviewed, some could not be found and those evaluations could therefore not be given a satisfactory rating on this criterion.

Figure 4.9 Cluster 4 Ratings: Comparison between Periods (first period n=9, second period n=13)



Finding 8: All evaluations demonstrated clear linkages between the evidence, findings, conclusions, lessons learned, and recommendations overall but cross-cutting issues were not well reflected in the findings and conclusions in almost half the reports reviewed spanning both periods. Moreover, lessons learned, where applicable, were either not included or not correctly identified in half of the reports, particularly those from the second period.

There are evident links between the findings, conclusions, and recommendations overall, despite in some cases the conclusions and recommendations not addressing the full range of key findings or conclusions simply summarizing some key findings without adding significant value to them (see Figure 4.9). The weaknesses most observed within this cluster of criteria and across both periods concern the assessment of cross-cutting issues and the formulation of lessons learned where these were applicable.

While such issues as gender equality, fragility, the environment, and inclusive growth are clearly addressed as being important in the Bank's ten-year strategy 2013-2022¹⁵, there appears to have been a missed opportunity to address these issues in a cross-cutting manner in the evaluations reviewed, even the second period evaluations that were conducted following the publication of the ten-year strategy. For example, while findings on cross-cutting issues may have been presented in evaluation reports, assessment of these was not always carried over to the conclusions, or lessons learned and recommendations where applicable. In some cases, cross-cutting issues were also variously identified in the reports such that it was not always very clear what exactly the cross-cutting issues were. Moreover, in some evaluations where certain cross-cutting themes would have been relevant to address, they were inexplicably overlooked. The impact evaluation is a case in point. While inclusive economic growth and gender equality were assessed in terms of the impact of water/WASH interventions on women's employment in Ethiopia, measured through a community survey, other cross-cutting issues such as related to the environment were not addressed yet are clearly pertinent to water sector interventions.

Concerning lessons learned, several corporate and country evaluations in particular, did not adequately address these. In cases where the evaluations did not include them and did not specifically aim to provide these, i.e. where lessons were not specified in the statement of purpose or objectives, this criterion was deemed non applicable which therefore did not affect the rating for that cluster. However, in cases where the evaluation specified the identification of lessons in the purpose or objectives statements yet none were identified in the summary report, the rating for this criterion was negatively affected. Ratings were also negatively affected where lessons were reported but they were not clearly or correctly worded to show broader relevance.

Case Example: The African Development Bank's Support for Agricultural Value Chains Development: Lessons for the Feed Africa Strategy

Both the summary and technical reports are well written, concise, and clearly structured. They include an executive summary that contains all the relevant information drawn from the evaluation. The evidence presented appears to come from multiple sources although it is not always clear from which groups of stakeholders (primary sources only). The evaluation criteria and questions are systematically answered. The findings thoroughly discuss cross-cutting issues of gender, inclusiveness, and environmental sustainability, although environmental issues could be given more attention in the recommendations. The conclusions are appropriately derived from the findings and provide additional insights but the lessons learned from the evaluation are scant. The recommendations are pertinent.

¹⁵ AfDB, *At the Center of Africa's Transformation: Strategy for 2013-2022*. 2013, pp. 1-3.

5 Conclusions, Lessons Learned, and Recommendations

5.1 Conclusions

Based on the results of this quality assessment, the quality of evaluations has improved since the implementation of IDEV's evaluation strategy. However, it is important to note that we cannot attribute the improvement to the strategy itself given the multiple factors that can influence quality. While country and thematic/sector evaluations showed the sharpest increase in "satisfactory" and "highly satisfactory" ratings between the two periods (almost 40% increase in the case of country evaluations), the quality of corporate evaluations, already significantly higher than country and thematic/sector evaluations in the 2009-12 period, increased to a lesser degree in the 2015-17 period.

The majority of quality assessment criteria were rated "satisfactory" and "highly satisfactory", particularly in the case of evaluations conducted in the second period. However, there remains room for improvement in some areas. Weaknesses were notably observed in relation to specifying the evaluation design and including a complete evaluation matrix with the methodology, describing stakeholders and their engagement in the evaluation process, explicitly addressing the full range of evaluation principles, effectively integrating cross-cutting issues, and formulating lessons learned.

Cross-cutting issues such as gender equality, inclusive growth, environment/climate change, and fragility that are explicitly mentioned as areas of focus in AfDB's ten-year strategy are not always clearly identified in the evaluations reviewed and where they are identified, their analysis lacks depth. The findings reveal uneven coverage of these dimensions both within each type of evaluation product and across the range of products. These dimensions are commonly referred to as "cross-cutting" yet they are not always effectively integrated in a cross-cutting manner in evaluations.

Similarly, there remains uneven coverage of evaluation principles among evaluations conducted in the second period. While quality assurance and knowledge sharing through the dissemination of findings were rated quite highly, principles with respect to impartiality, ethics, coordination and alignment, partnership and capacity development were not explicitly addressed in the evaluation products reviewed. Key stakeholder engagement is also often not clearly described. While these principles may be considered in practice, they are not adequately documented.

With respect to evaluation methodology, the figures presented indicate that the evaluation design is not always clearly described nor the evaluation matrix adequately presented.

Findings reported appear to be evidence-based and supported by extensive secondary data but limited primary sources. Some overlap was noted between information in the different sections of the reports,

for example findings including evaluator judgments and performance ratings, conclusions essentially repeating some key findings, and lessons learned worded like conclusions, even recommendations. Similarly, with respect to process, some overlap was also noted between some evaluation products, e.g. approach papers and ToR.

The 6-point rating scale typically used by IDEV for the evaluations in the sample, was cumbersome and overly specific, conducive to a micro-level analysis that is neither practical nor particularly relevant, given the rather broad scope of evaluations conducted by IDEV, both in terms of geographical coverage and timeframe considered. While this scale has recently been replaced by a 4-point scale, the new scale is also problematic to a certain extent because its labeling does not identify a point in the range between “unsatisfactory” and “satisfactory”.

5.2 Lessons Learned

The following lessons have been drawn from this quality assessment:

Lesson 1: The testing of data collection tools is a necessary step in the assessment/evaluative process, both to ensure that tools are appropriately adapted to the evaluative situation and that all data collectors use the tools consistently. This facilitates subsequent data compilation and analysis and strengthens the validity of findings.

Lesson 2: A team-based approach as opposed to single, individual assessment of evaluations strengthens consistency in ratings and evaluative judgment.

Lesson 3: Involving key stakeholders in the discussion and/or development of recommendations strengthens ownership and the likelihood that recommendations will be effectively implemented.

Lesson 4: Imposing strict limitations on the maximum number of pages for evaluation reports (in this case summary reports), especially in complex evaluations such as those produced by IDEV, can result in some important elements necessary for a good understanding of the evaluation being omitted. This can impede full reader understanding of the evaluation process and issues and affect the quality of evaluation reporting.

Lesson 5: Having clear guidelines on the basic elements that should be included in each evaluation product – and each of the main sections of the different types of product – clarifies requirements for reporting and contributes to ensuring more even quality of documents produced.

5.3 Recommendations

Based on the findings and conclusions of this quality assessment, we put forth some recommendations related to the different evaluation products for IDEV-AfDB in their support of the evaluation function.

Recommendation 1: Consider adding a requirement for stakeholder mapping. This requirement would be stipulated in the initial evaluation planning documents and the stakeholder map developed by the evaluators and presented in both the inception and evaluation technical reports (in the annexes if

too detailed). This would ensure all key stakeholders (internal and external) are identified and their roles and contributions clearly described.

Recommendation 2: IDEV should define what cross-cutting themes to prioritize on the basis of what AfDB policy has identified as important and should identify which of these themes are relevant for each type of evaluation:

- Relevant cross-cutting themes should be explicitly identified at the outset (inception phase) and effectively integrated in a cross-cutting manner, beginning with the description of the context and object of the evaluation through the evaluation objectives, scope, findings, conclusions, and recommendations;
- Where the absence of disaggregated data on these themes impedes their full assessment, the evaluation can be an opportunity to probe these issues further through their integration in the evaluation framework and corresponding primary data collection instruments and to recommend to management that more attention be paid to them in the monitoring of Bank operations.
- With respect to gender equality in particular, while it may not be deemed a relevant cross-cutting theme in all types of evaluations conducted by IDEV, given that the AfDB, like all international financial institutions, has a policy that defines its commitment “to promote gender mainstreaming as a means of fostering poverty reduction, economic development and gender equality on the continent recognizing gender equality as a development objective”¹⁶, and because IDEV led the drafting of guidance for ECG on gender-responsive evaluations and integrating gender analysis into project-level evaluation¹⁷, IDEV should attempt to systematically integrate gender considerations into all evaluations that involve the assessment of projects as well as other types of evaluations where it would be relevant to do so.
- Where support is needed for the effective integration of these issues, IDEV should seek expert technical assistance through specialized workshops, webinars, etc. aimed at building the capacities of task managers in these areas.

Recommendation 3: Ensure that OECD-DAC and AfDB-IDEV evaluation principles related to evaluation independence, impartiality, ethics, partnership, coordination, and capacity building are clearly and explicitly stated and explained in the various evaluation products, from the design to final report stage, and considered as part and parcel of the evaluative process. Evidence of consideration of these principles could be provided following the description of the evaluation methodology, in the introduction of various data collection tools, and in a disclaimer in the opening pages of evaluation reports (both summary and technical reports).

Recommendation 4: Acknowledging IDEV’s recent decision to replace the cumbersome 6-point rating scale with a 4-point scale for future evaluations, consider re-labeling the new scale to allow for a more nuanced assessment between “unsatisfactory” and “satisfactory” where this would be relevant or applicable. The scale should allow for a perspective that more positively reflects some degree of progress

¹⁶ African Development Bank-African Development Fund. June 2001. *The Gender Policy*. OESU, p. i.

¹⁷ See Evaluation Cooperation Group. 2017. *Integrating Gender into Project-level Evaluation*. Abidjan: IDEV-AfDB.

made toward satisfactory performance. We recommend replacing the existing rating labels with the following: “unsatisfactory”, “partly unsatisfactory”, “satisfactory” and “highly satisfactory”.

Recommendation 5: Clearly identify the basic elements that each evaluation product should contain and clarify the circumstances under which ToR, approach papers, and inception reports should be produced.

Recommendation 6: Develop guidance for evaluators with clear definitions of key evaluation terms to ensure a distinction is made between the type of information that should be included in the different sections of the evaluation report, e.g. evaluation findings, conclusions, lessons learned, and recommendations.

Recommendation 7: Ensure that recommendations are discussed systematically with the relevant operations sections of the Bank and clearly prioritized or classified with a corresponding timeframe to help inform decision making. While avoiding the formulation of recommendations that are too prescriptive, IDEV should identify issues of concern within the recommendations to connect them more specifically to the evaluation findings reported.

Recommendation 8: Ensure the annexes in the inception report include the data collection tools and the annexes in the summary report contain an abridged evaluation matrix, a key evaluation tool that links the scope of the evaluation to the evaluation methodology. The abridged matrix could include the evaluation criteria and main corresponding questions as well as a sample of key indicators – both quantitative and qualitative where relevant – and the data collection methods for each main question.

Appendix I List of Evaluations Reviewed

YEAR	EVALUATION TITLE	TYPOLGY
2009	Cameroun: Évaluation de l'assistance de la Banque	Country
2009	Cap Vert: Évaluation de l'assistance de la Banque 1996-2007	Country
2009	Independent Evaluation of the Decentralization Strategy and Process at the African Development Bank	Corporate
2010	Evaluation of Policy-based Operations in the African Development Bank, 1999-2009	Thematic/sector
2010	Independent Assessment of the Quality at Entry of ADF 2005-2008 Operations and Strategies	Corporate
2010	Project Supervision at the African Development Bank 2011-2008	Corporate
2010	Unlocking the Potential of Africa's Sub-Regions: Review of Bank Group Assistance to Sub-Regional Development Banks	Thematic/sector
2011	Gabon Country Assistance Evaluation, 1996-2008	Country
2012	Mainstreaming Gender Equality: A Road to Results or a Road to Nowhere?	Others: Synthesis
2015	Evaluation of Bank Assistance to Small and Medium Enterprises (2006–2013)	Thematic/sector
2015	Independent Evaluation of Policy and Strategy Making and Implementation	Corporate
2016	Comprehensive Evaluation of the Development Results of the African Development Bank Group 2004-2013	Corporate
2016	Democratic Republic of Congo - Evaluation of the Bank's Country Strategy and Program 2004-2015	Country
2016	Ghana: Evaluation of the Bank's Country Strategy and Program 2002–2015	Country
2016	Impact Evaluation of the Rural Water Supply and Sanitation Program in Ethiopia 2006-2014	Others: Impact
2016	South Africa - Country Strategy and Program Evaluation 2004-2015	Country
2016	Towards Private Sector Led Growth: Lessons of Experience - Evaluation Synthesis Report	Others: Synthesis
2017	Evaluation des stratégies et programmes de la Banque Africaine de Développement BAD en Côte d'Ivoire 2006-2016	Country
2017	Evaluation of the Regional Integration Strategy and Operations of the African Development Bank, 2011-2016	Others: Regional
2017	Spurring local socio-economic development through rural electrification	Others: Cluster
2017	The African Development Bank's Human Resource Management Policy and Strategic Directions: A Formative Evaluation	Corporate
2017	The African Development Bank's Support for Agricultural Value Chains Development: Lessons for the Feed Africa Strategy	Thematic/sector

Appendix II Quality Assessment Template

Evaluation Information	
Title of evaluation	
Type of evaluation	
Quality assessors	
Evaluation delivery year	
Document(s) available and reviewed	

CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
i) Context, object, purpose, objectives, and scope of the evaluation		
<p>1. The background/context of the object of the evaluation is clearly described and includes all relevant factors.</p> <p><i>Key social, cultural, political, economic, and institutional factors that are relevant to the object of the evaluation.</i></p>		
<p>2. A detailed description of the object of the evaluation is included.</p> <p><i>Clear identification of the object of the evaluation, its geographic location, timeframe, goal, objectives, stakeholders, beneficiaries, budget, implementation status, key components/activities, expected results.</i></p>		
<p>3. The purpose of the evaluation, including its intended use and users and the relevance of its timing, are clearly presented.</p> <p><i>Clear statement describing the evaluation purpose, its rationale, intended use, intended users, and the relevance of its timing.</i></p>		
<p>4. The specific objectives of the evaluation are clearly outlined and</p>		

CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
coherent with the purpose. <i>Relevant listing of specific evaluation objectives that are well aligned with the evaluation purpose statement.</i>		
5. The scope of the evaluation is clear and coherent with the purpose of the evaluation. <i>Clear identification of geographic areas and timeframe covered, and programmatic inclusions/ exclusions. The scope is aligned with the evaluation purpose.</i>		
6. Evaluation criteria are identified and coherent with the purpose of the evaluation. <i>Reference to standard OECD-DAC evaluation criteria, namely: relevance, effectiveness, efficiency, and sustainability. Other criteria, if used, should be explained.</i>		
7. Cross-cutting issues are sufficiently addressed in the description of the context, evaluation object and scope, where applicable. <i>Sufficient attention paid to relevant cross-cutting issues, such as gender equality, environment/climate, inclusive growth, fragility, etc., in the description of the context, evaluation object and scope, where applicable.</i>		
OVERALL ASSESSMENT SECTION I <i>Explanation of the average section rating, summarizing the strengths and weaknesses identified.</i>	Average rating section i	
ii) Design and methodology		
8. The evaluation design is clearly described and appropriately aligned with the evaluation purpose. In the case of theory based evaluations, a		

CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
<p>clear and coherent theory of change of what is being evaluated is provided which includes assumptions and is used to inform the evaluation questions.</p> <p><i>Clear and relevant information on evaluation design, such as experimental, quasi-experimental, non-experimental, theory-based, etc. If theory based, clear and coherent description of the logic model/theory of change of the object of the evaluation. This includes a description of the assumptions underlying the causal linkages and of the indicators/ judgment criteria/ benchmarks used for measuring results achievement.</i></p>		
<p>9. The report clearly identifies the main sources of information and adequately explains data collection, sampling, and analysis methods.</p> <p><i>Clear identification of primary (where applicable) and secondary sources of information and explanation of: quantitative and/or qualitative data collection methods, such as online surveys, semi-structured interviews, focus groups, etc.; sample, including selection criteria for each method; quantitative and/or qualitative data analysis methods, such as triangulation, statistical analysis, most significant change, etc.</i></p>		

CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
<p>10. Evaluation questions are well aligned with the evaluation criteria identified and supported by clear and measurable indicators or judgment criteria.</p> <p><i>Evaluation questions are linked to the evaluation criteria and presented in an evaluation matrix that should comprise at least 4 columns (criteria/questions, sub-questions, indicators, and data collection methods for each question) and should be included in annex if not in main report.</i></p>		
<p>11. The evaluation risks and limitations, and appropriate mitigation strategies for each, are clearly explained.</p> <p><i>Clear description of evaluation risks and limitations faced and mitigation strategies used. Any significant differences from the planned approach should be explained.</i></p>		
<p>OVERALL ASSESSMENT SECTION II</p> <p><i>Explanation of the average section rating, summarizing the strengths and weaknesses identified.</i></p>	<p>Average rating section ii</p>	
iii) Process		
<p>12. There are logical links between information presented in key evaluation products, e.g. ToR, approach paper, inception report, and summary evaluation report.</p> <p><i>Consistency in the evaluation objectives, scope, design and methodology with each successive product building upon the previous. Any differences in approach introduced during the process should be explained.</i></p>		

CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
<p>13. The report identifies the key stakeholders that were engaged in the evaluation process and explains how they were engaged.</p> <p><i>Identification of key stakeholders consulted for the evaluation, their level of consultation, roles and responsibilities of each, including of a steering or other advisory committee if any.</i></p>		
<p>14. The evaluation quality assurance process is clearly described.</p> <p><i>Clear information on how the quality was/would be ensured in the evaluation, e.g. stakeholder involvement in validation of findings, peer review of evaluation report, etc.</i></p>		
<p>15. There is evidence to indicate that the evaluation findings were/would be disseminated/communicated.</p> <p><i>Knowledge-sharing plan in approach paper and/or evidence that key evaluation findings were disseminated/communicated through a debriefing, workshop, publication, etc.</i></p>		
<p>16. There is evidence of consideration of other OECD-DAC and AfDB-IDEV evaluation principles, namely independence, impartiality, ethics, coordination and alignment, partnership and capacity development.</p> <p><i>Evidence of: organizational independence and avoidance of conflict of interest; avoidance of bias in findings reported; ethical guidelines and codes of conduct followed; national and local evaluation plans, activities and policies taken into account; and evaluation learning supported and</i></p>		

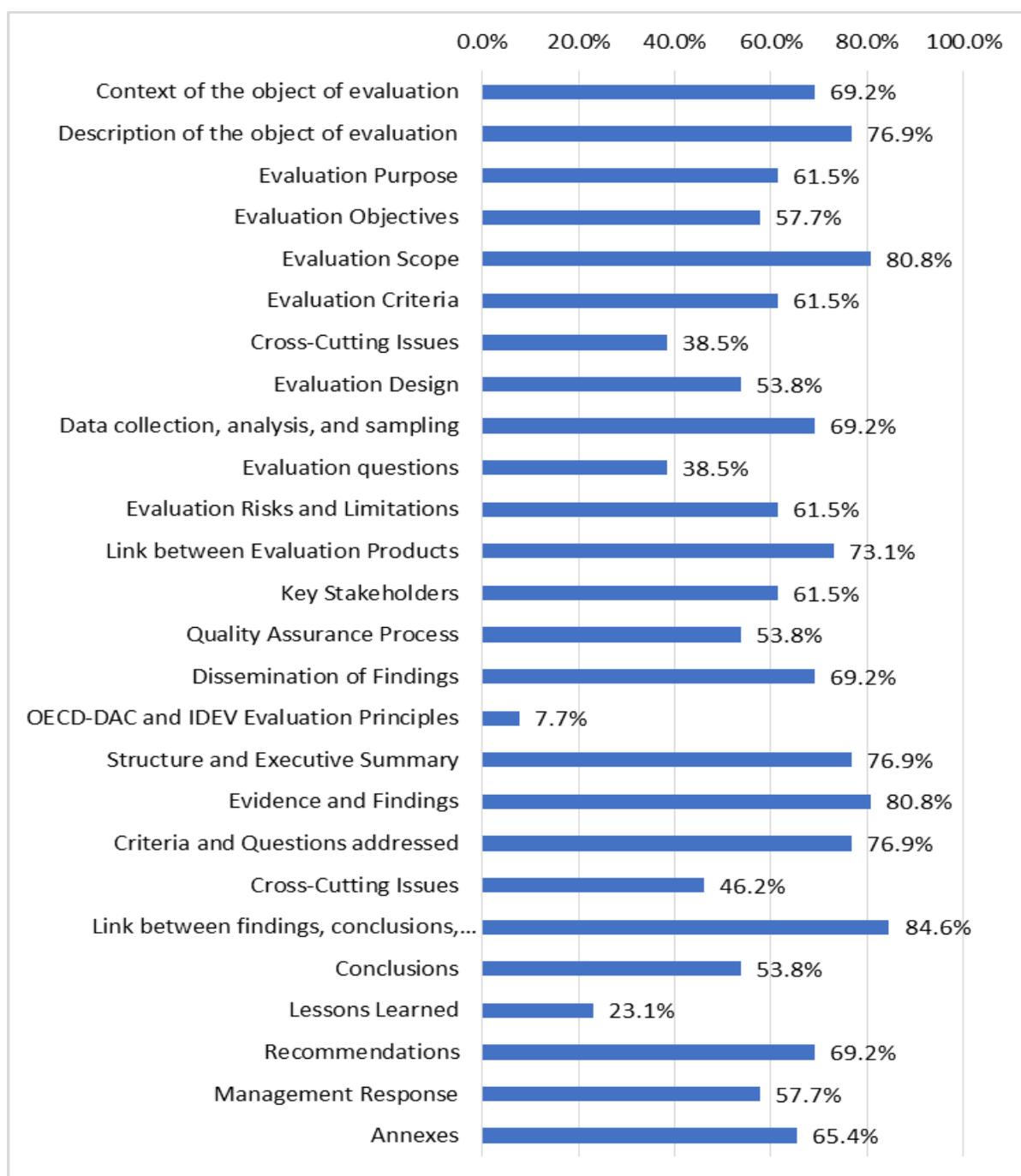
CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
<i>strengthened in regional members.</i>		
OVERALL ASSESSMENT SECTION III <i>Explanation of the average section rating, summarizing the strengths and weaknesses identified.</i>	Average rating section iii	
iv) Clarity and robustness of reporting		
17. The evaluation report has a clear and logical structure and concise, representative executive summary. <i>Logical order of presentation: background/context, object of the evaluation, evaluation purpose, objectives, scope, methodology, findings, etc. Succinct executive summary covering all key elements of the evaluation, e.g. maximum 5 pages.</i>		
18. The evaluation report presents clear analysis and findings are based on multiple (2+) lines of evidence. <i>Evidence of triangulation - at least 2 lines of evidence presented in findings – and diversity of perspectives. Evidence that data gaps, if any, are discussed and, where relevant, caveats are provided to guide the reader on the interpretation of findings.</i>		
19. The evaluation criteria and questions are systematically answered. <i>Findings should systematically respond to the evaluation criteria and questions and explanations provided for questions it was not possible to answer; any revisions to original questions are also documented.</i>		
20. Cross-cutting issues are sufficiently assessed in the findings and conclusions, and recommendations (where applicable). <i>Assessment of cross-cutting issues,</i>		

CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
<p><i>such as gender equality, environment/climate, inclusive growth, fragility, etc. across the evaluation findings and conclusions, and recommendations (where applicable).</i></p>		
<p>21. The links between the evidence, the findings and conclusions, lessons learned (if any), and the recommendations are evident.</p> <p><i>Coherence among the findings, conclusions, lessons learned (if any), and the recommendations. Conclusions, lessons, and recommendations should not introduce new information not addressed in the findings.</i></p>		
<p>22. Conclusions add value to the findings.</p> <p><i>Conclusions that provide evaluator insights based on the findings with respect to the main evaluation criteria, e.g. they describe the foreseeable implications of the findings for future initiatives. Conclusions should include a description of the main strengths and weaknesses of the object of the evaluation but should not simply repeat or summarize key findings.</i></p>		
<p>23. Lessons learned are clearly and correctly identified.</p> <p><i>Lessons that are clearly stated and demonstrate broader relevance/ applicability beyond the object of the evaluation itself.</i></p>		
<p>24. Recommendations are clear and specific, but not prescriptive, and indicate a timeline.</p> <p><i>Recommendations that are organized by order of priority and actionable but not prescriptive.</i></p>		
<p>25. There is evidence of an adequate management response to the evaluation recommendations (criterion not applicable to synthesis and cluster evaluations)</p>		

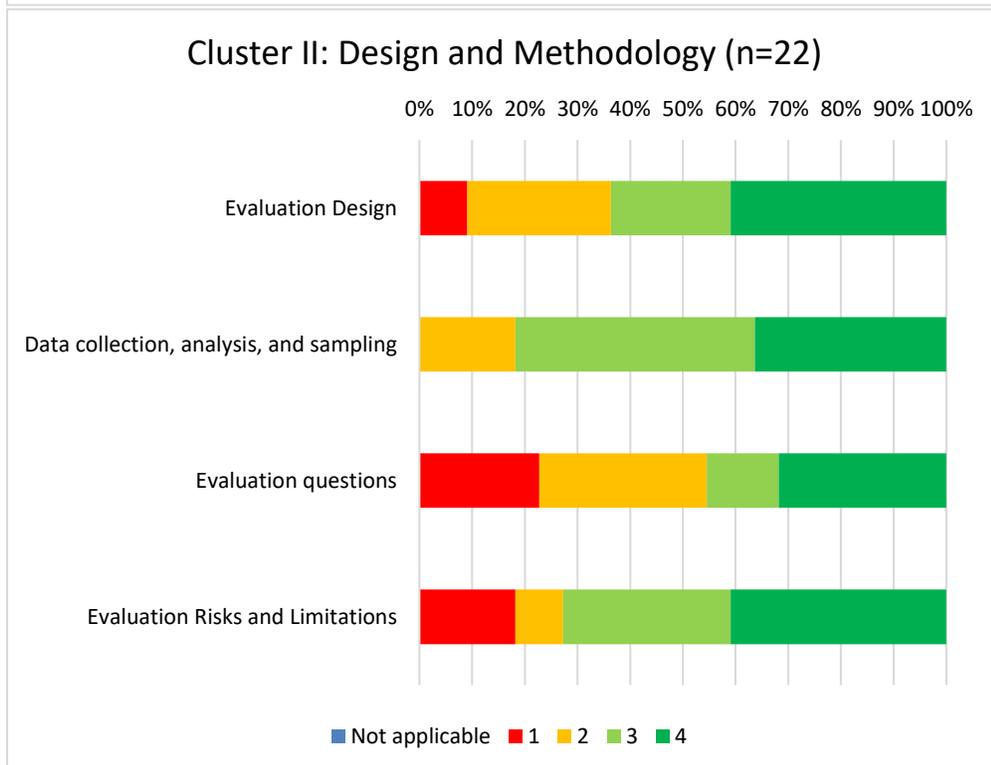
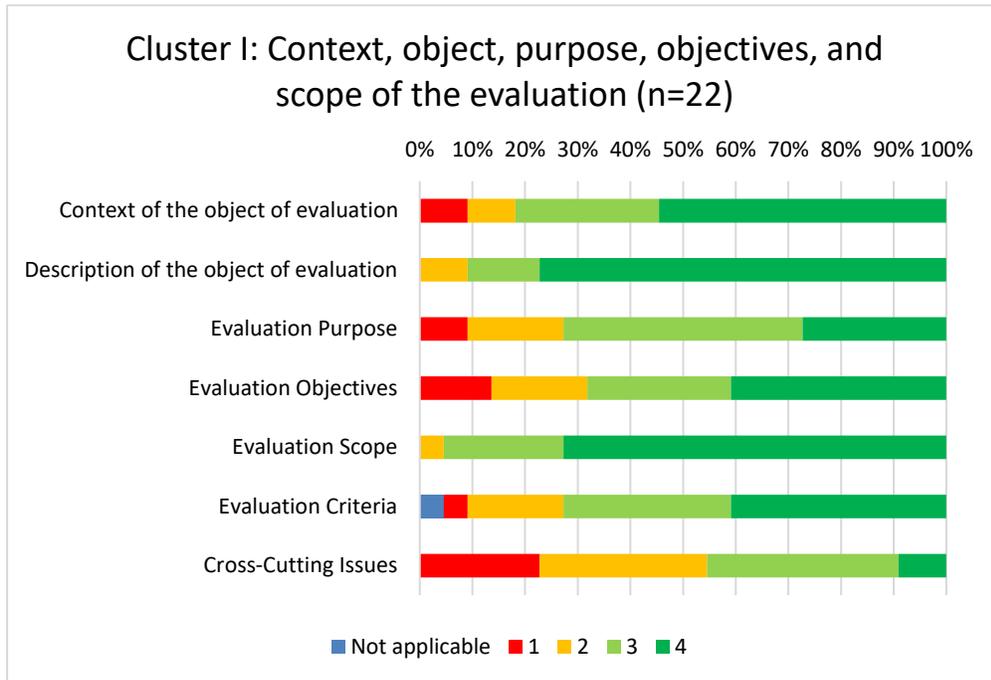
CRITERIA/SUB-CRITERIA	RATING	ASSESSOR COMMENTS*
<i>Type of management response, e.g. acceptance of most or all recommendations and timeline for implementation, explanation for rejecting any recommendations.</i>		
26. The report (detailed technical and/or summary) contains relevant annexes. <i>Complementary information, such as evaluation matrix, list of stakeholders consulted, other sources of information (provided this does not conflict with privacy and confidentiality), data collection tools, and data tables/additional evidence or context to support the analysis and findings presented, etc.</i>		
OVERALL ASSESSMENT SECTION IV <i>Explanation of the average section rating, summarizing the strengths and weaknesses identified.</i>	Average rating section iv	
OVERALL ASSESSMENT OF THE EVALUATION <i>Summary of the main strengths and weaknesses of the evaluation.</i>	Percentage of criteria rated 3 and 4	

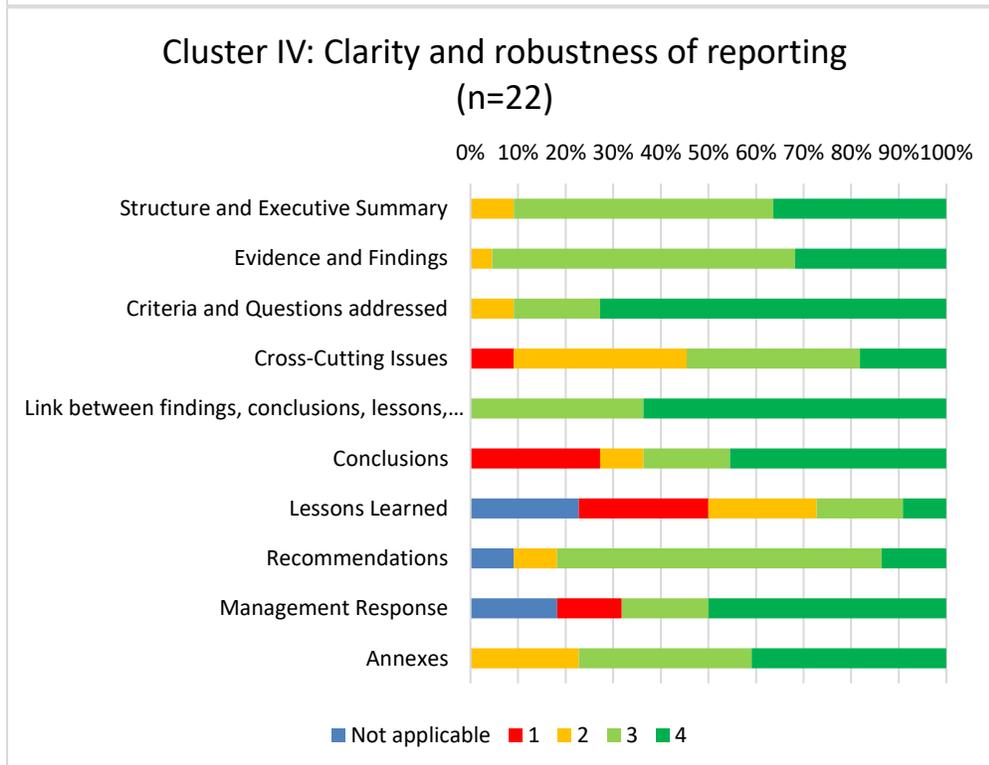
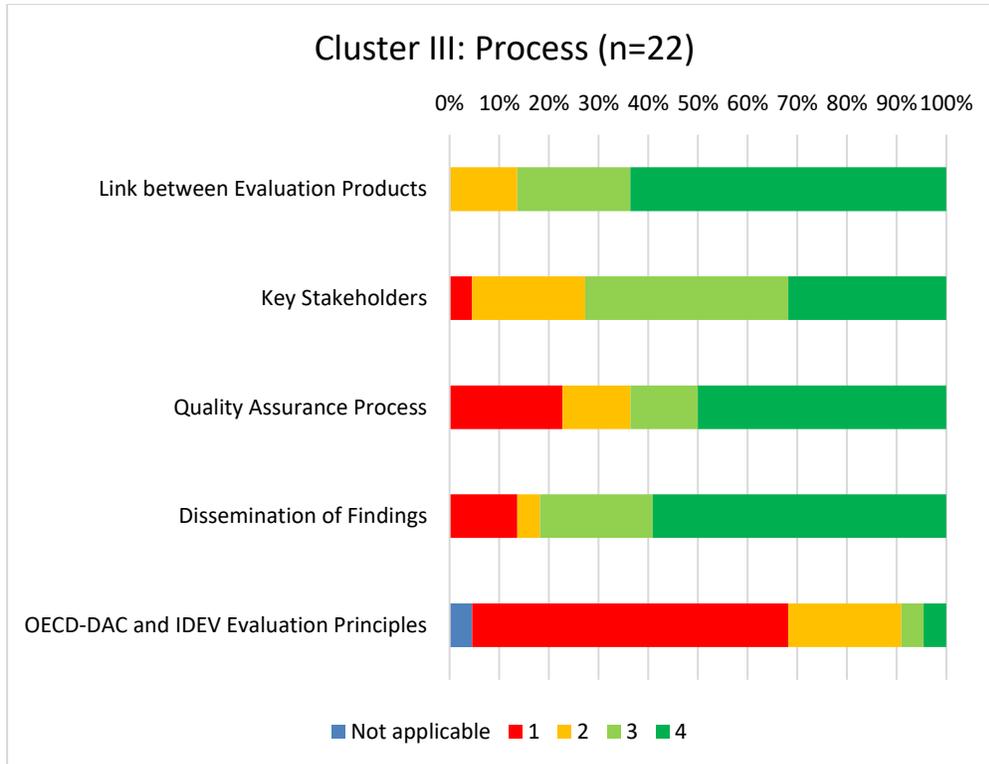
*Assessor comments will be based primarily on information found in the summary report of the evaluation but all associated documents will be reviewed, such as detailed technical reports and inception reports if available, to verify consistency in methodology for example but also to look for information that may not be found in the summary report. Where the information was not found in the summary report but in other associated documents, or not found in any documents reviewed, this will be explicitly mentioned.

Appendix III Percentage of criteria rated Satisfactory and Highly Satisfactory, All Evaluations

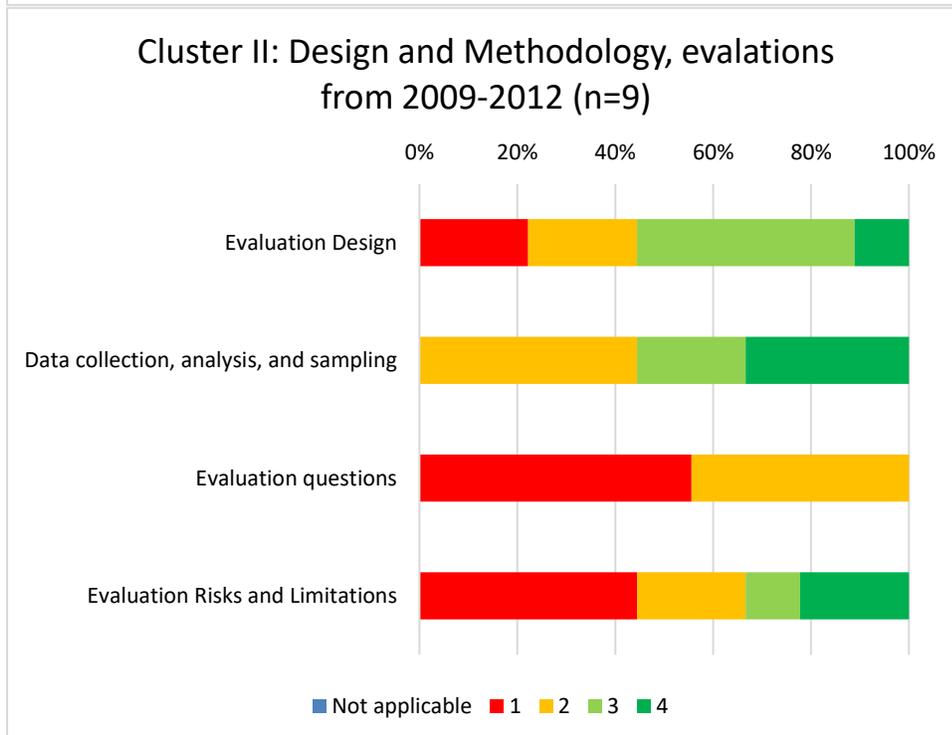
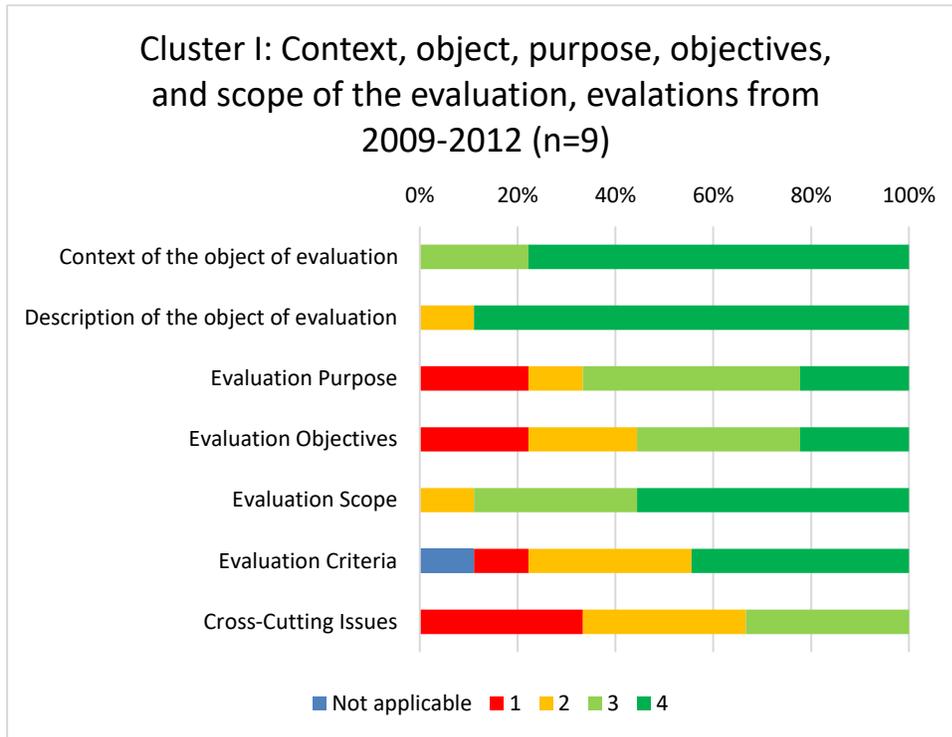


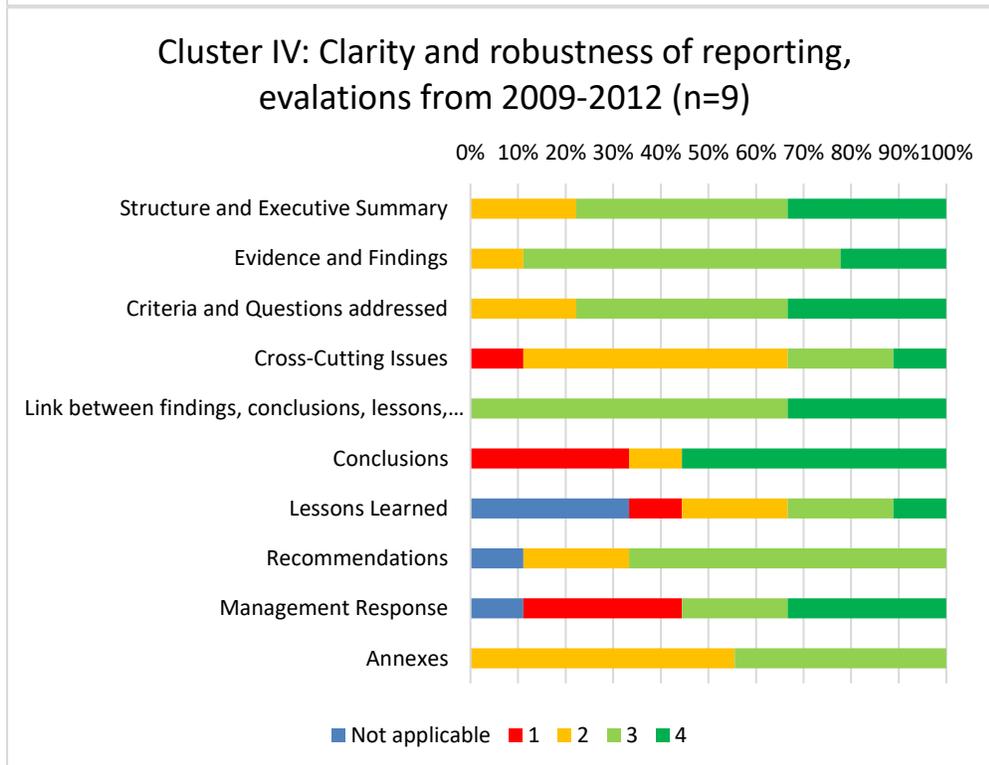
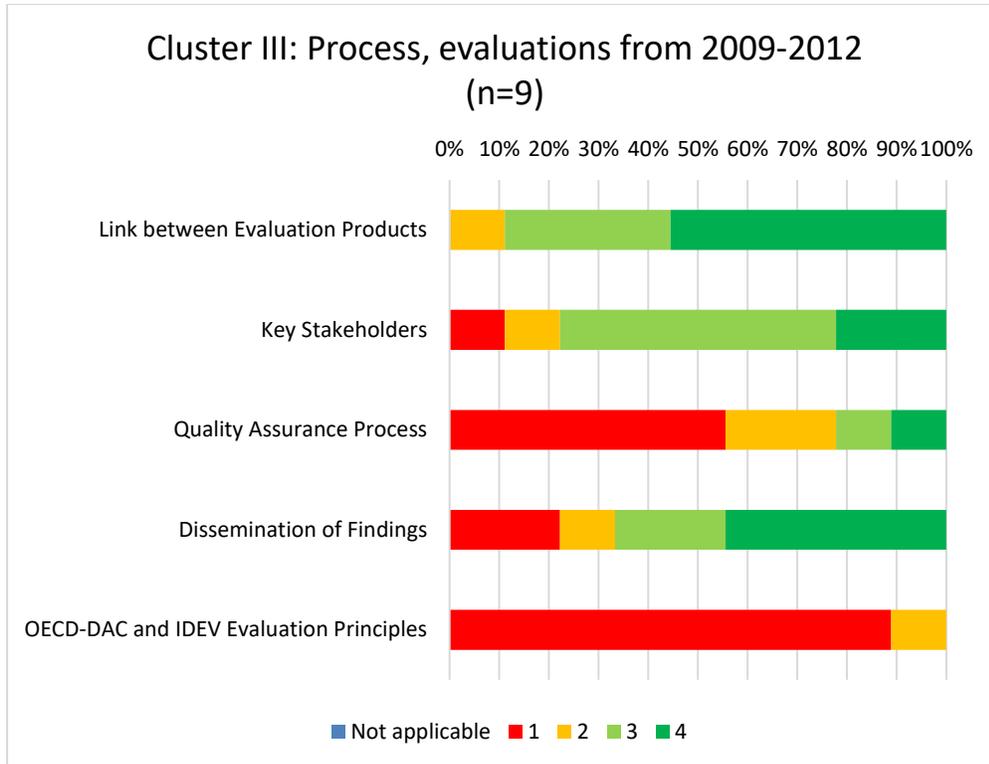
Appendix IV Quality Assessment Results by criteria, All Evaluations





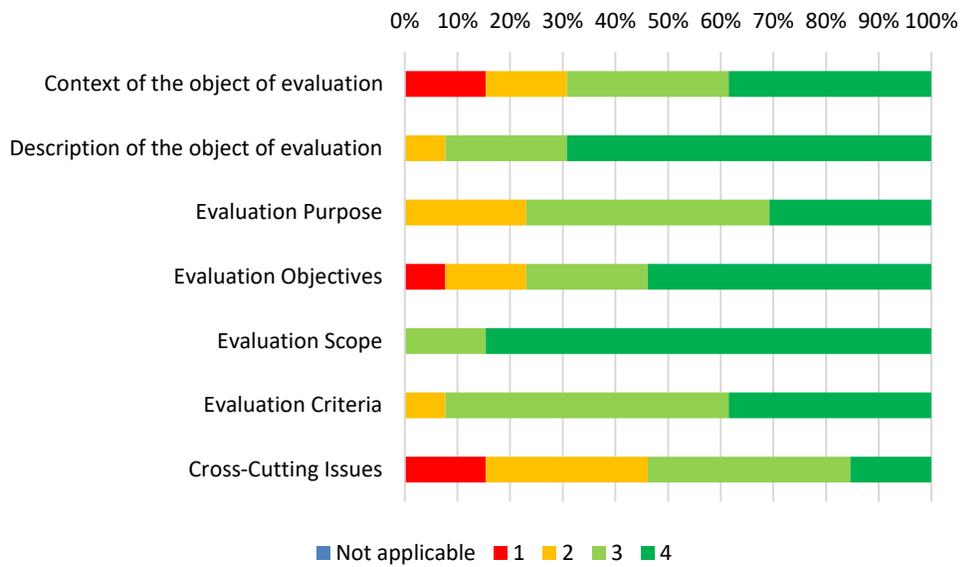
Appendix V Quality Assessment Results by criteria, First Period Evaluations (2009-2012)



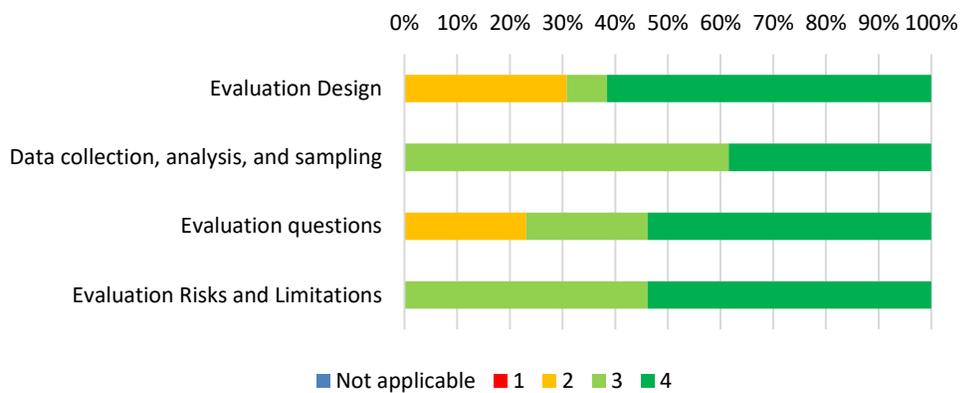


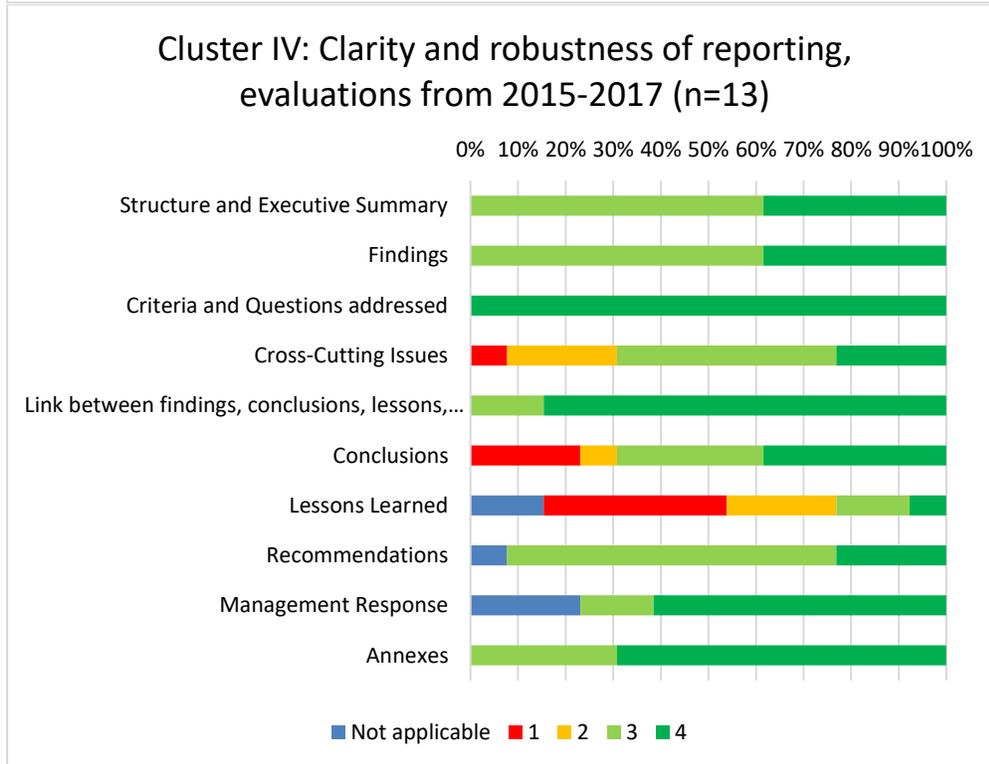
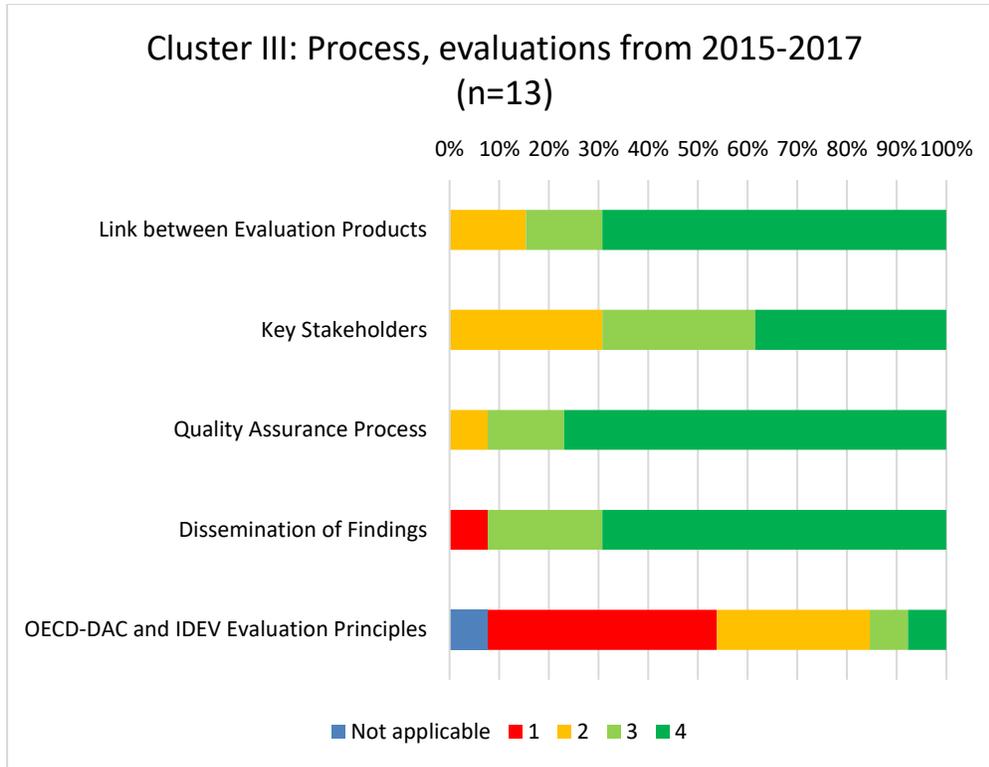
Appendix VI Quality Assessment Results by criteria, Second Period Evaluations (2015-2017)

Cluster I: Context, object, purpose, objectives, and scope of the evaluation, evaluations from 2015-2017 (n=13)

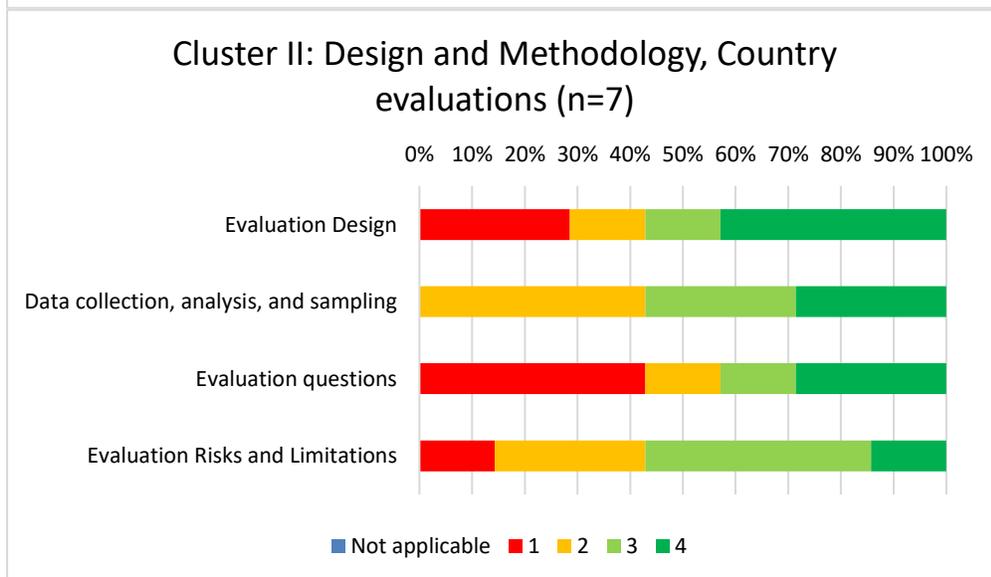
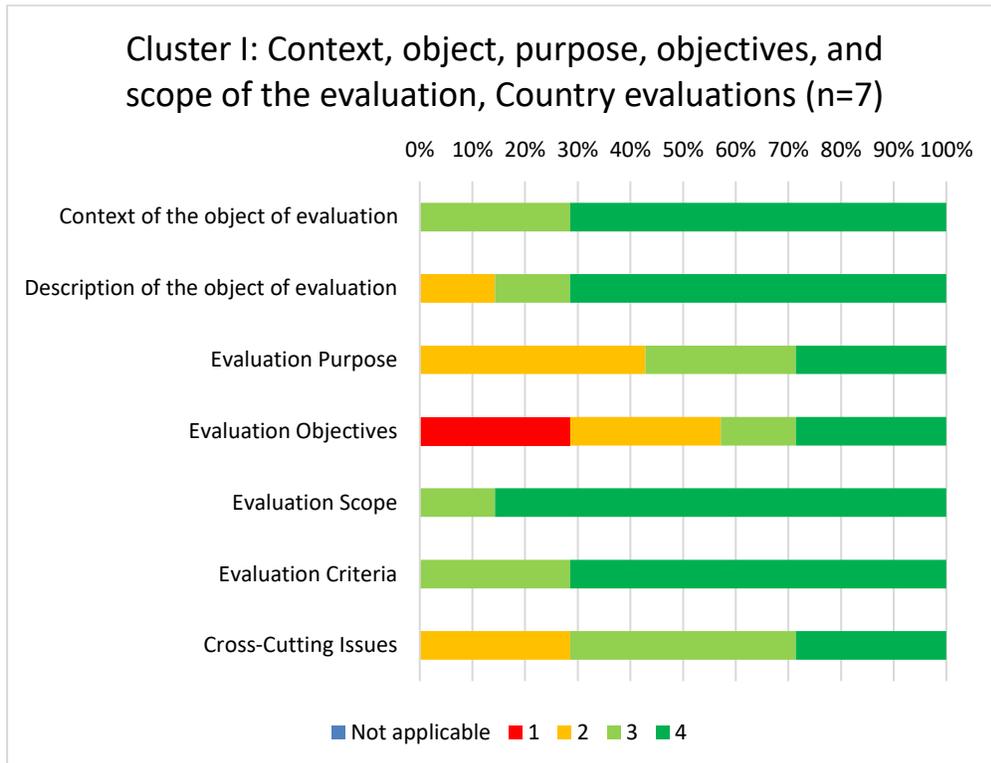


Cluster II: Design and Methodology, evaluations from 2015-2017 (n=13)

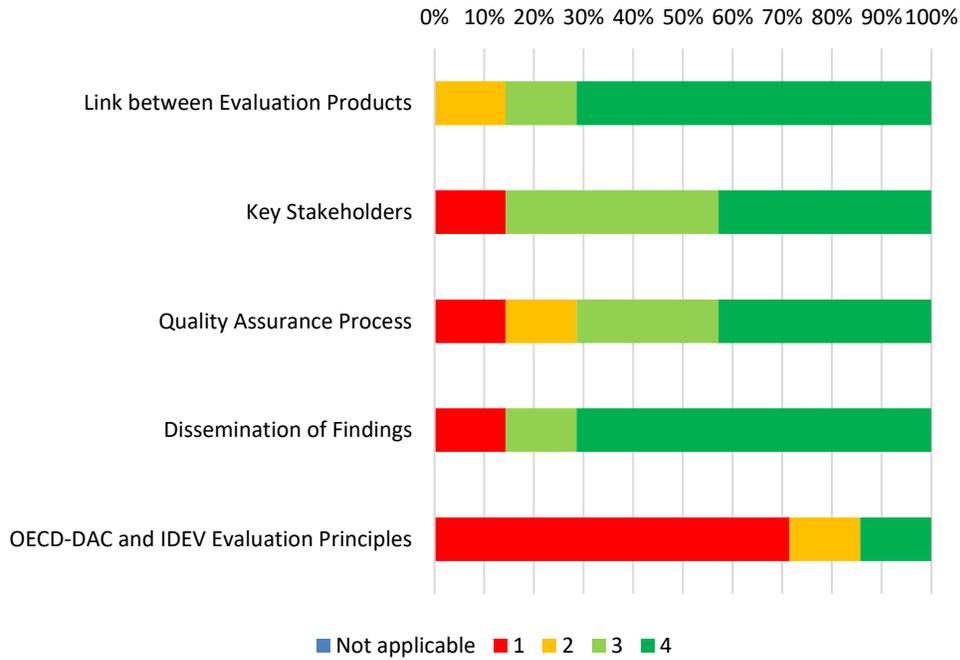




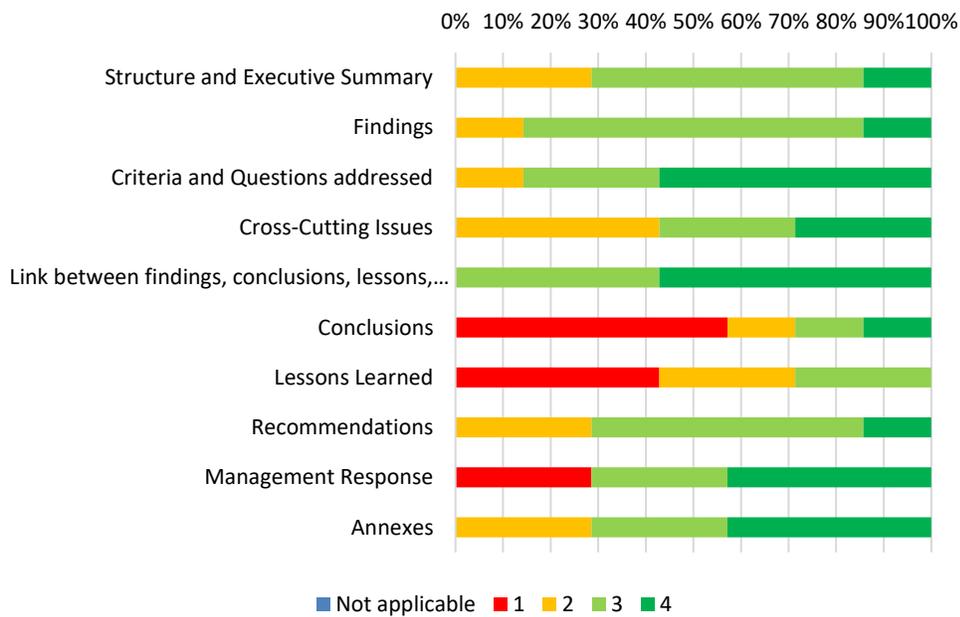
Appendix VII Quality Assessment Results by criteria, Country evaluations



Cluster III: Process, Country evaluations (n=7)

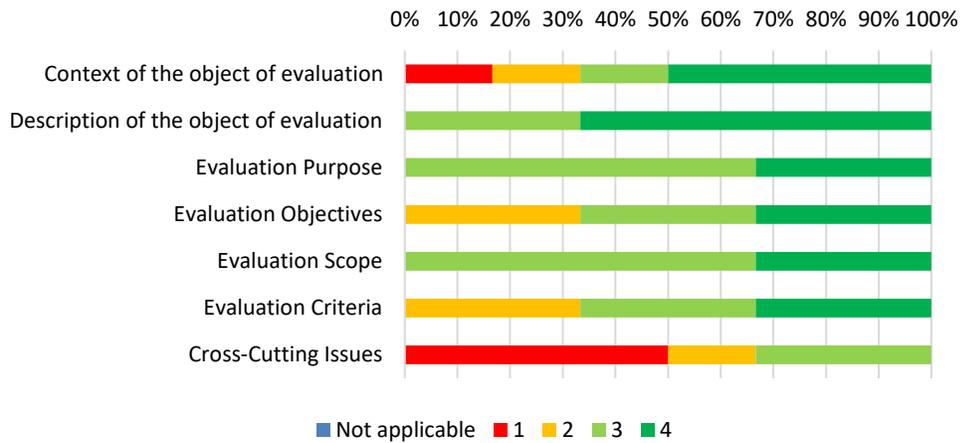


Cluster IV: Clarity and robustness of reporting, , Country evaluations (n=7)

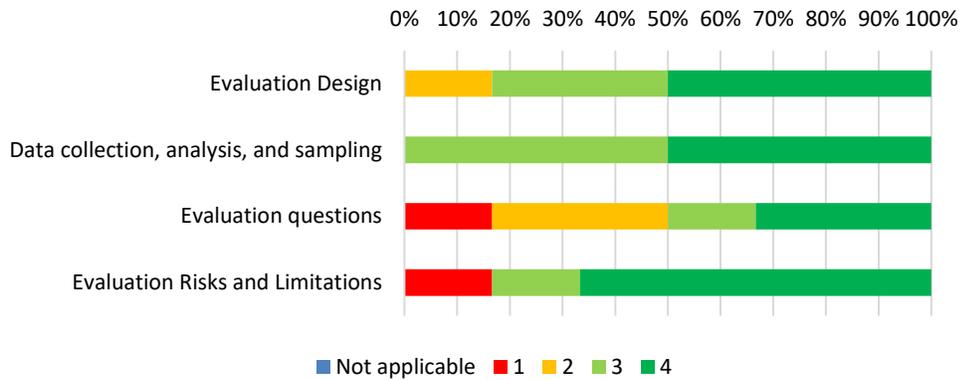


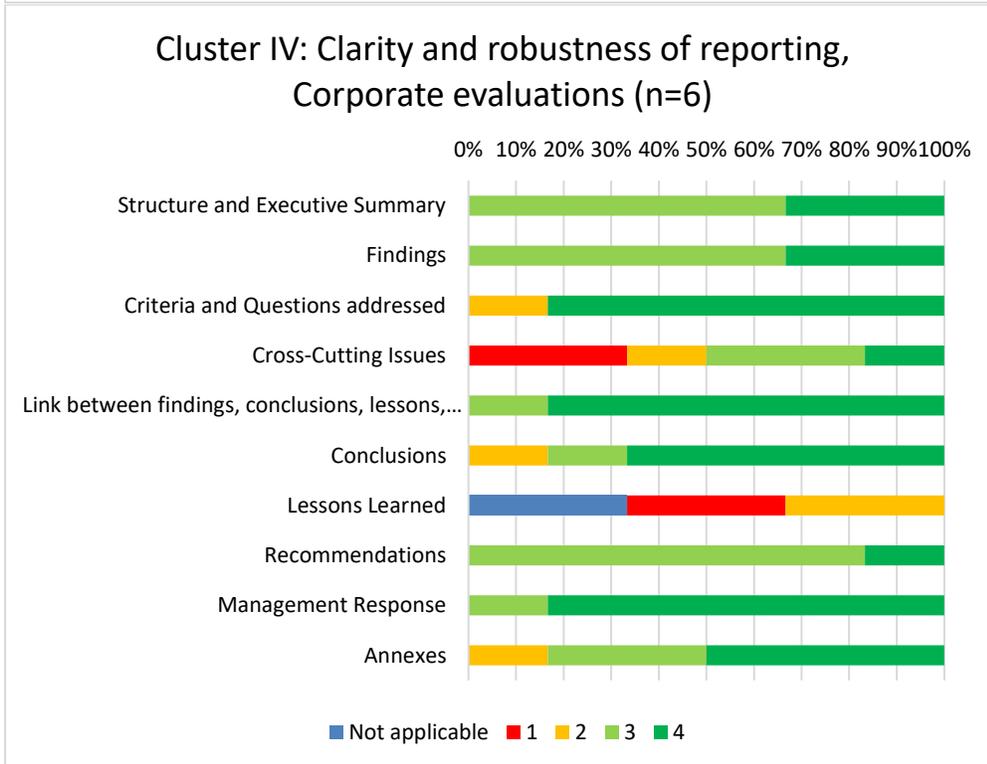
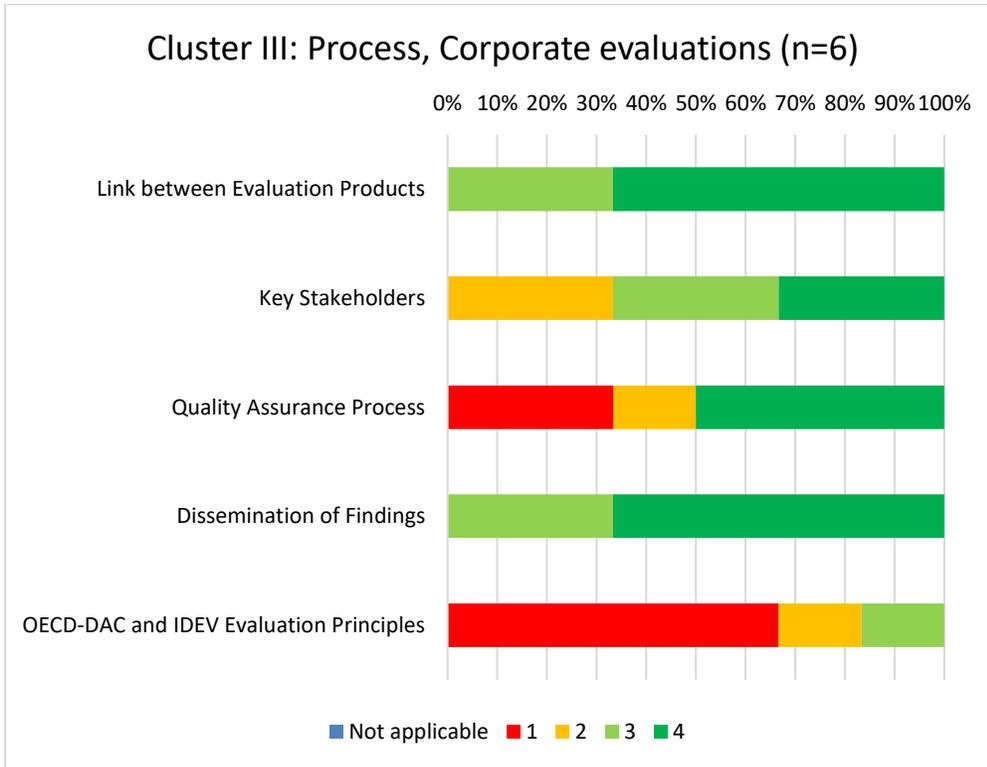
Appendix VIII Quality Assessment Results by criteria, Corporate Evaluations

Cluster I: Context, object, purpose, objectives, and scope of the evaluation, Corporate evaluations (n=6)



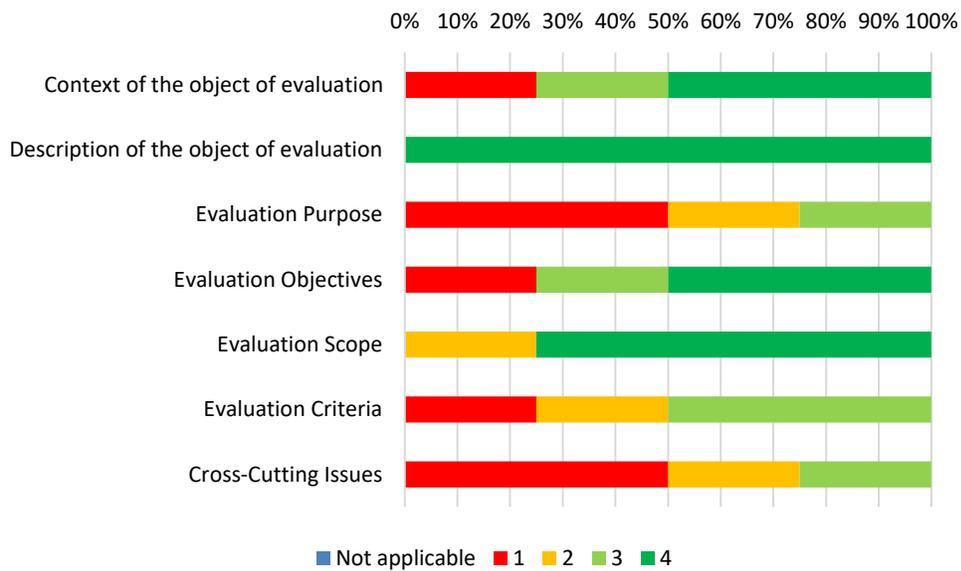
Cluster II: Design and Methodology, Corporate evaluations (n=6)



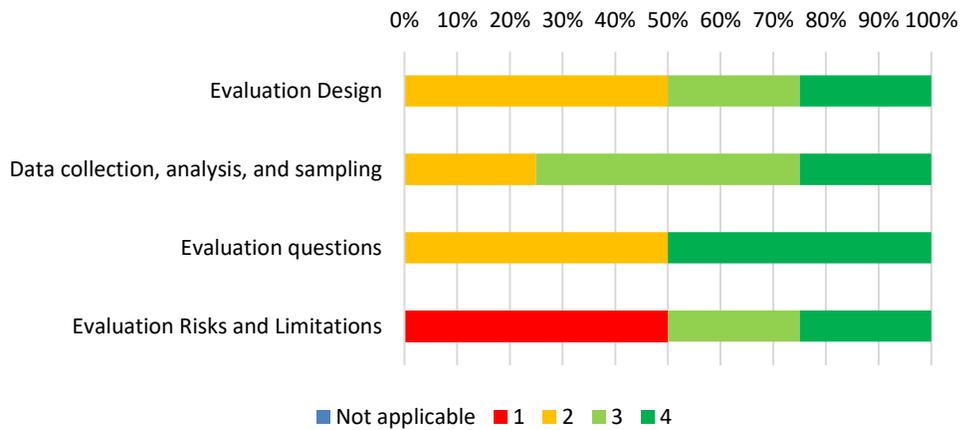


Appendix IX Quality Assessment Results by criteria, Thematic and Sector Evaluations

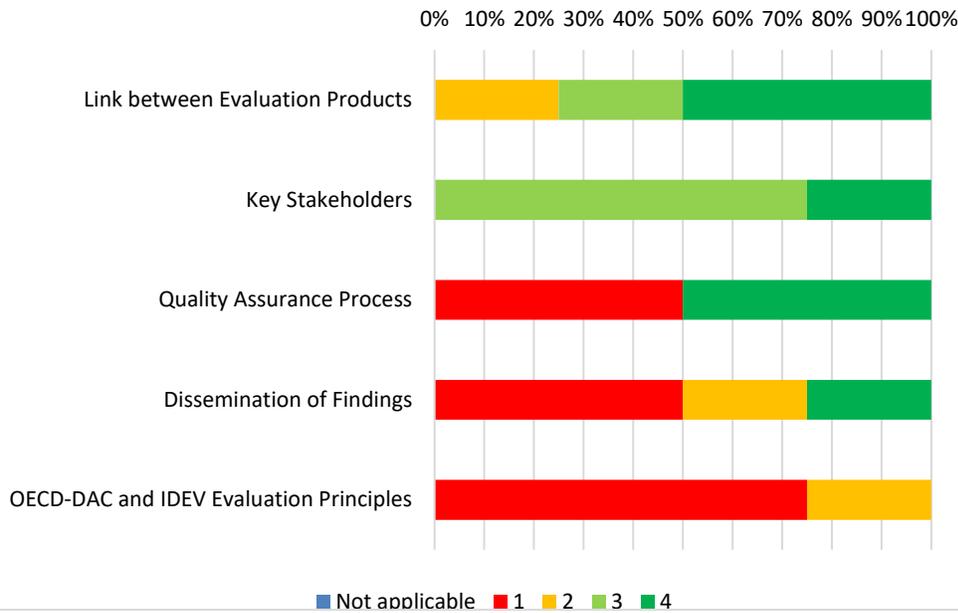
Cluster I: Context, object, purpose, objectives, and scope of the evaluation, Thematic and Sector evaluations (n=4)



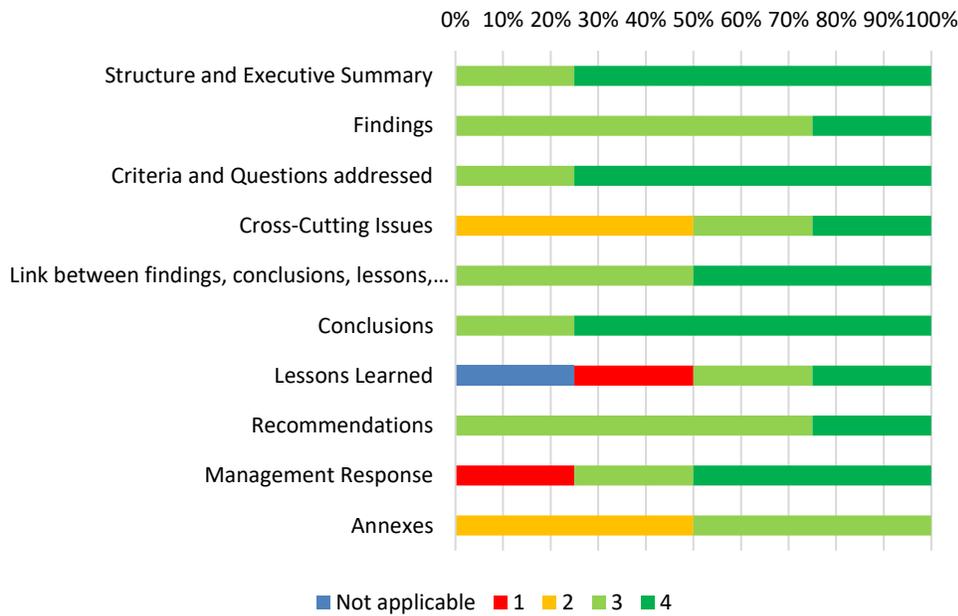
Cluster II: Design and Methodology, Thematic and Sector evaluations (n=4)



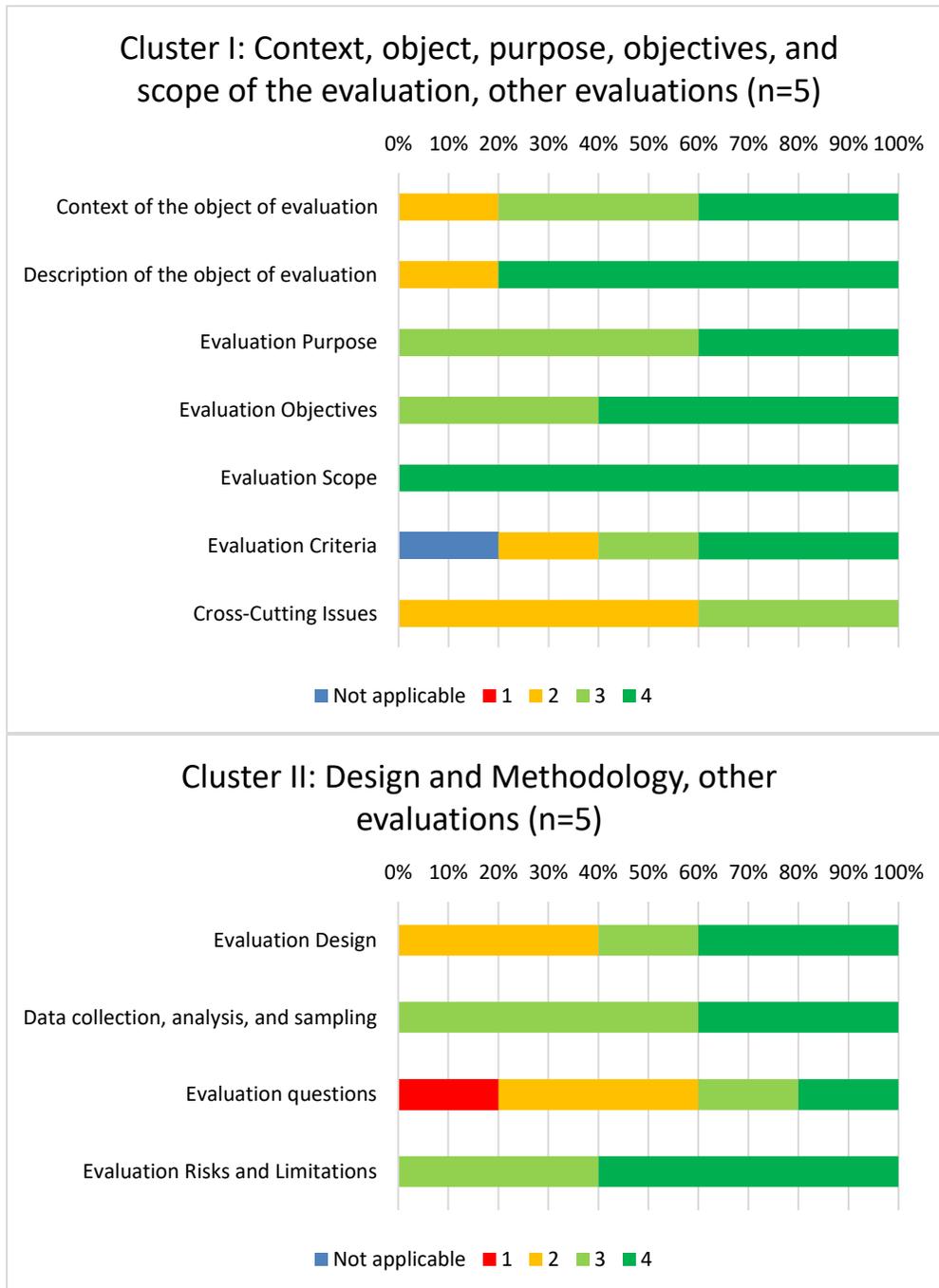
Cluster III: Process, Thematic and Sector evaluations (n=4)



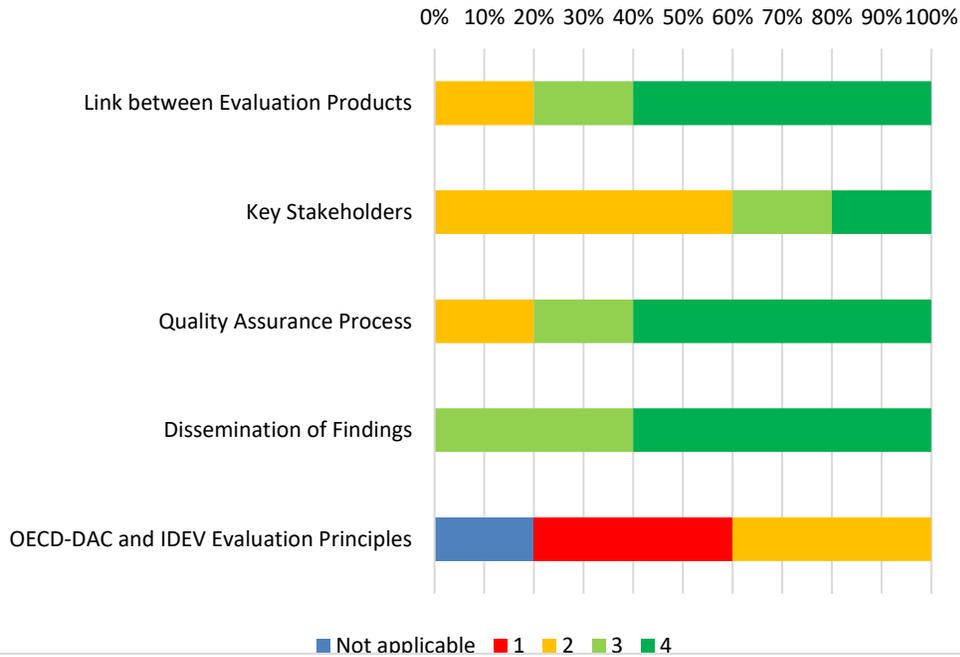
Cluster IV: Clarity and robustness of reporting, Thematic and Sector evaluations (n=4)



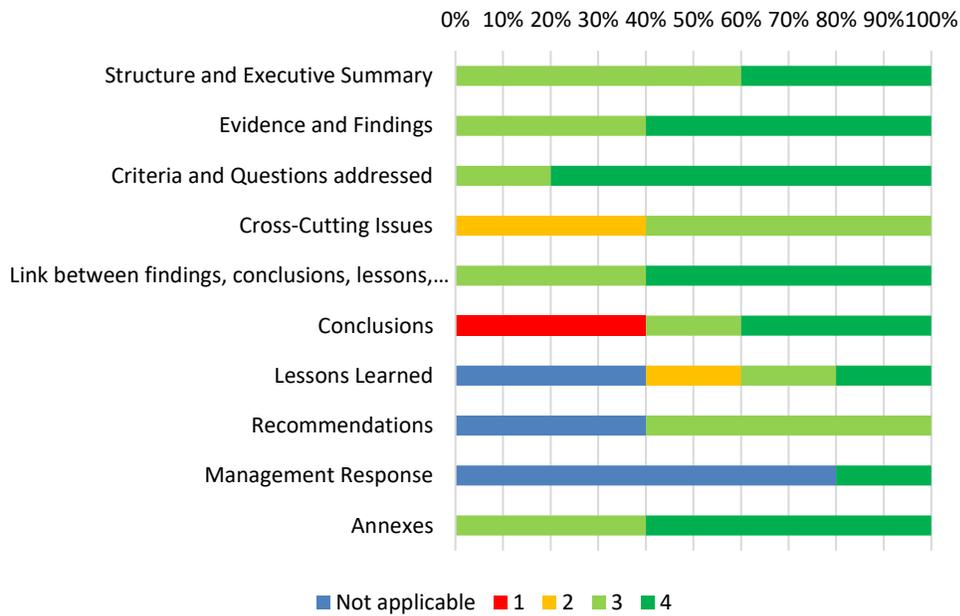
Appendix X Quality Assessment Results by Criteria, Other Evaluations



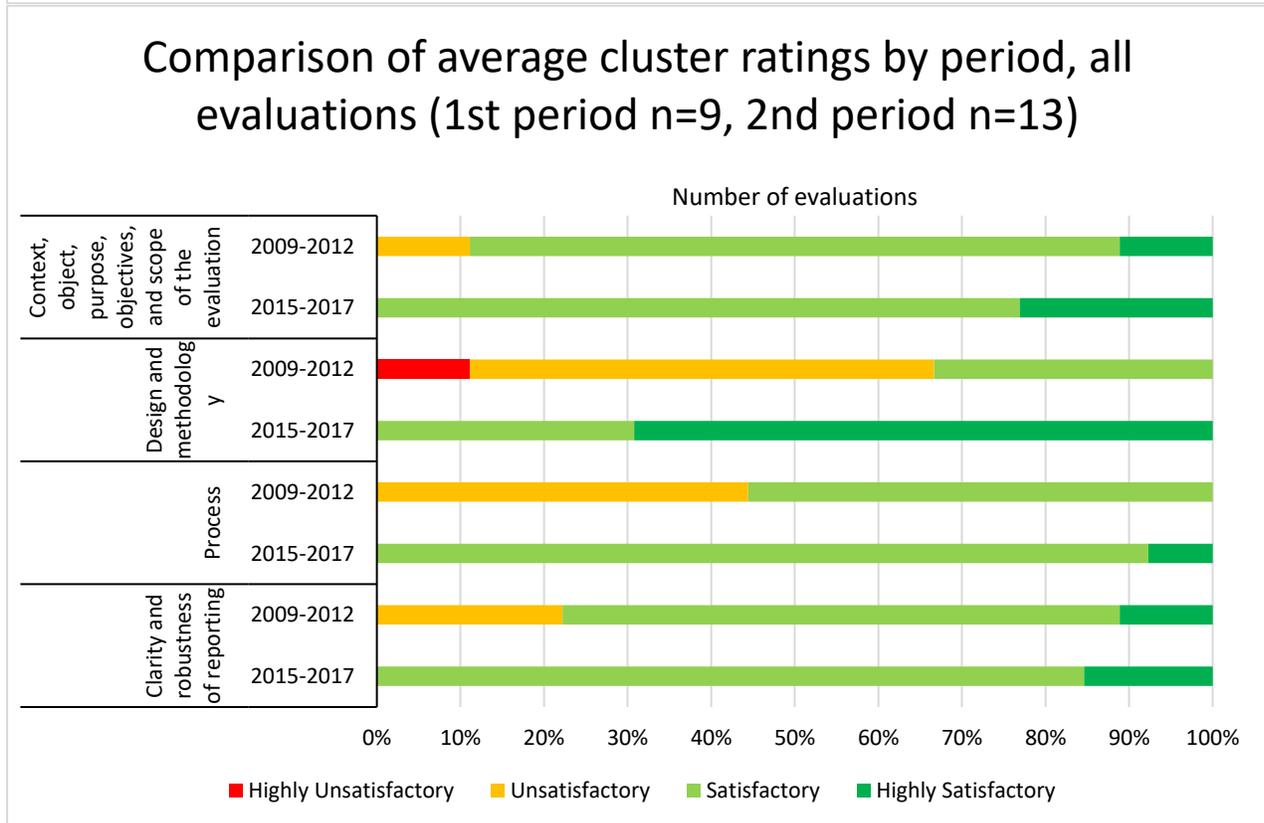
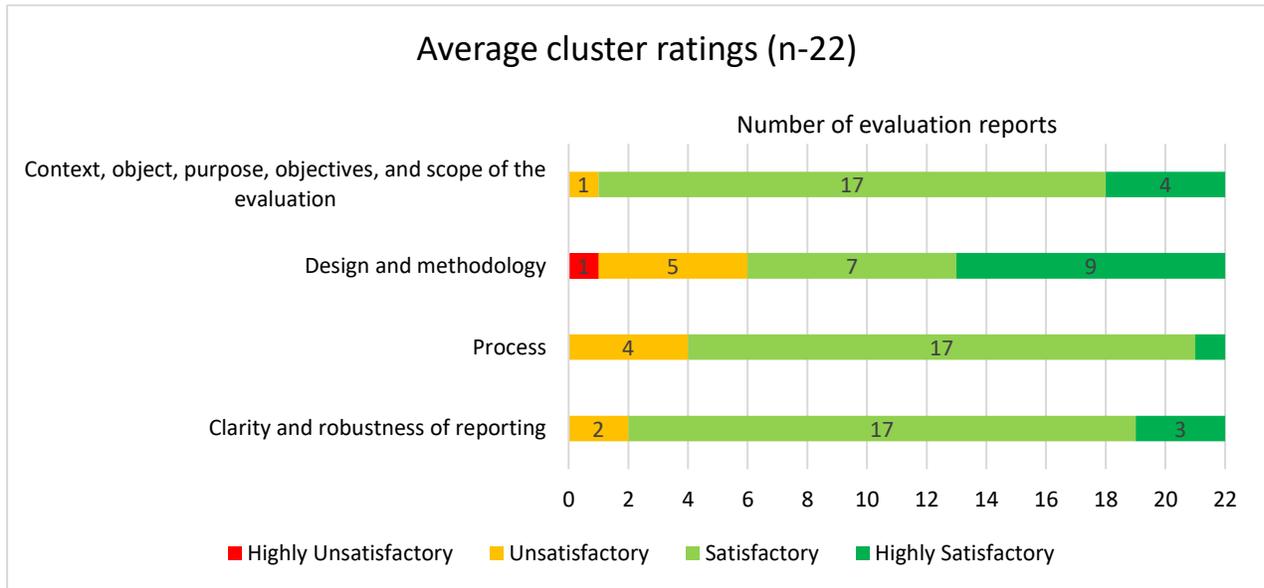
Cluster III: Process, other evaluations (n=5)



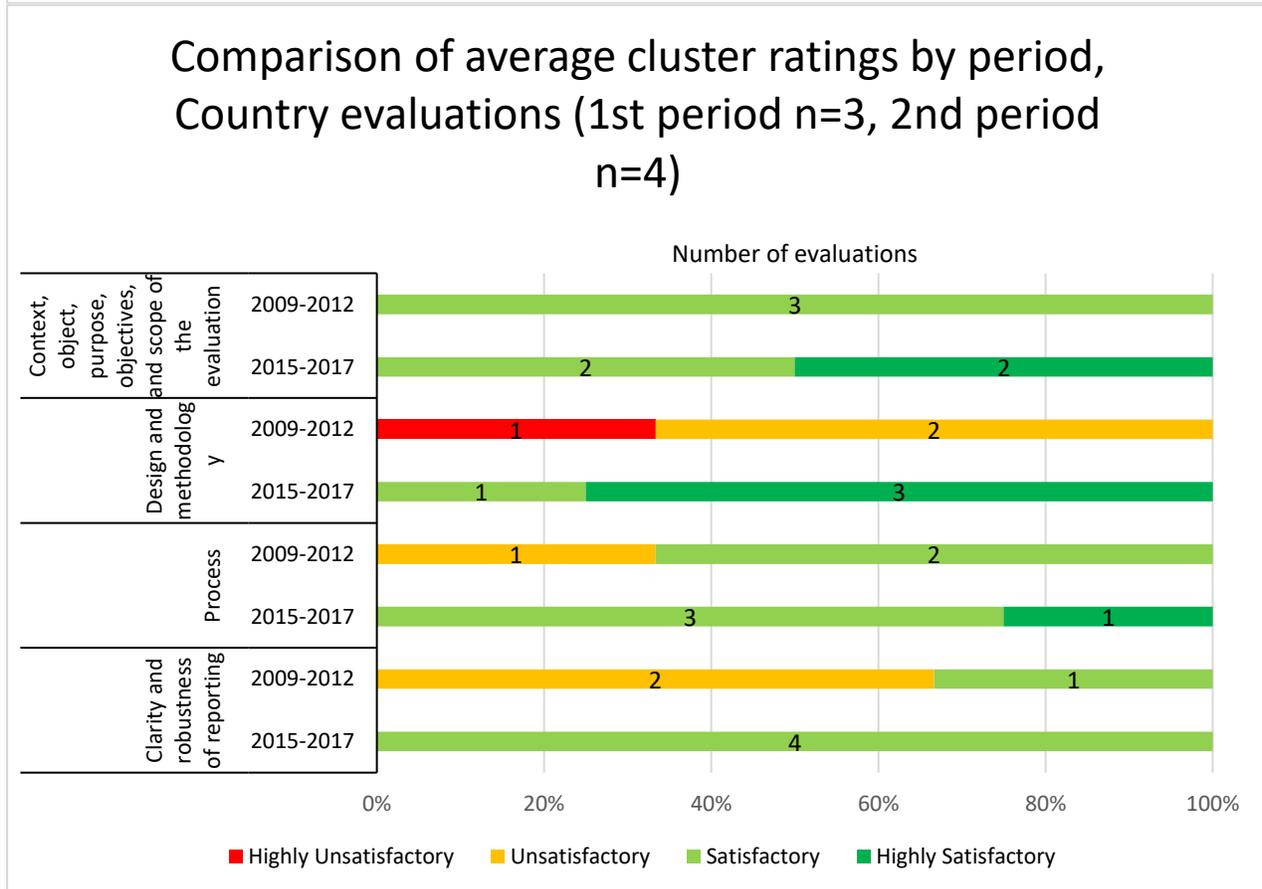
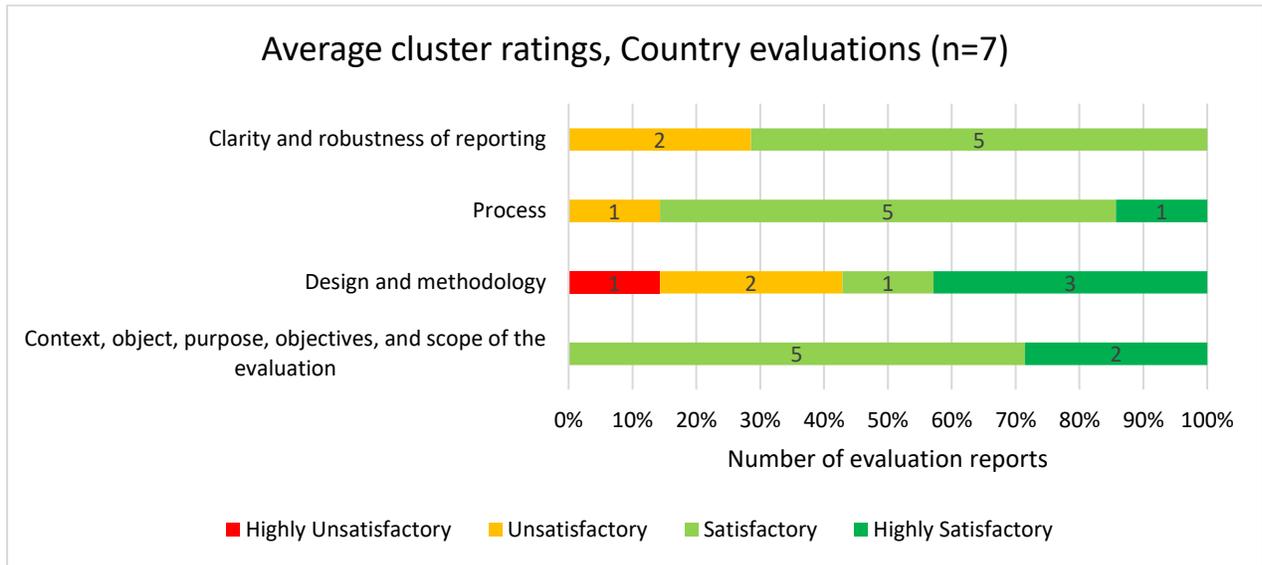
Cluster IV: Clarity and robustness of reporting, other evaluations (n=5)



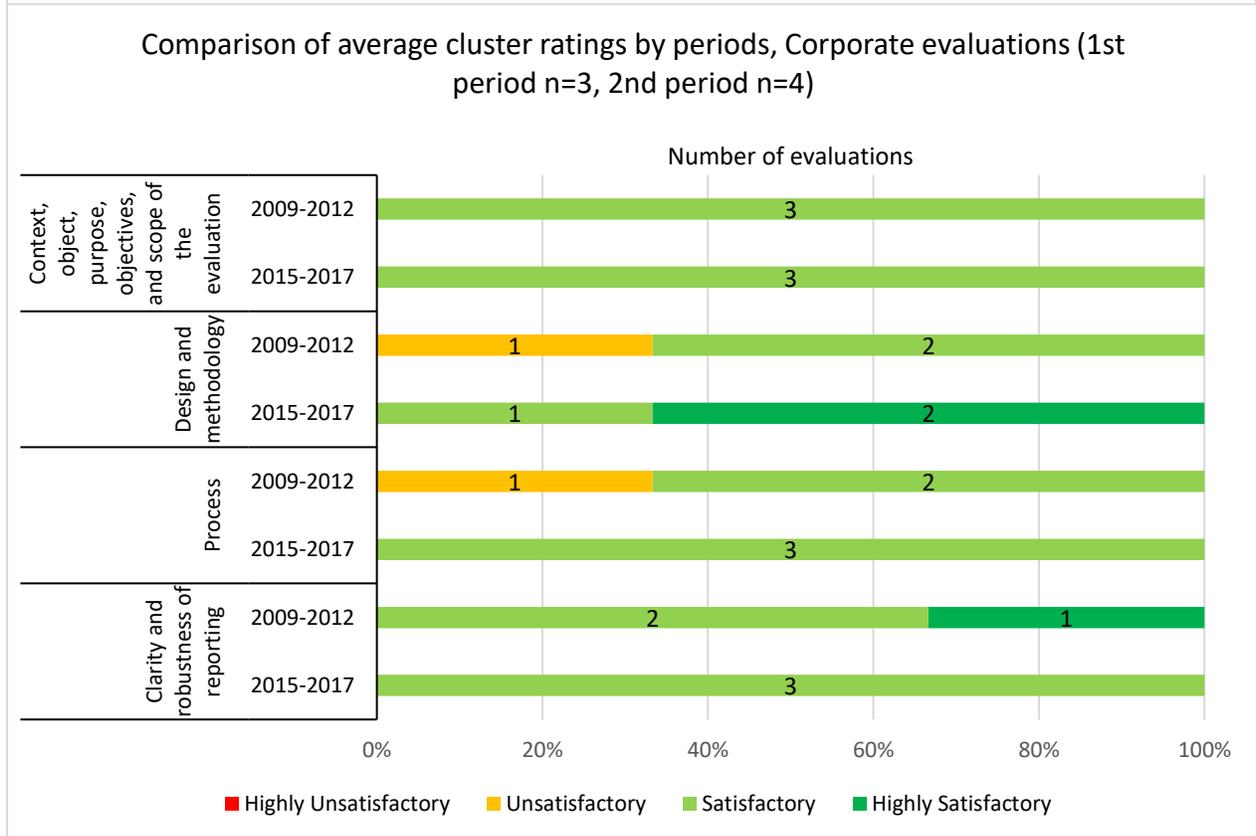
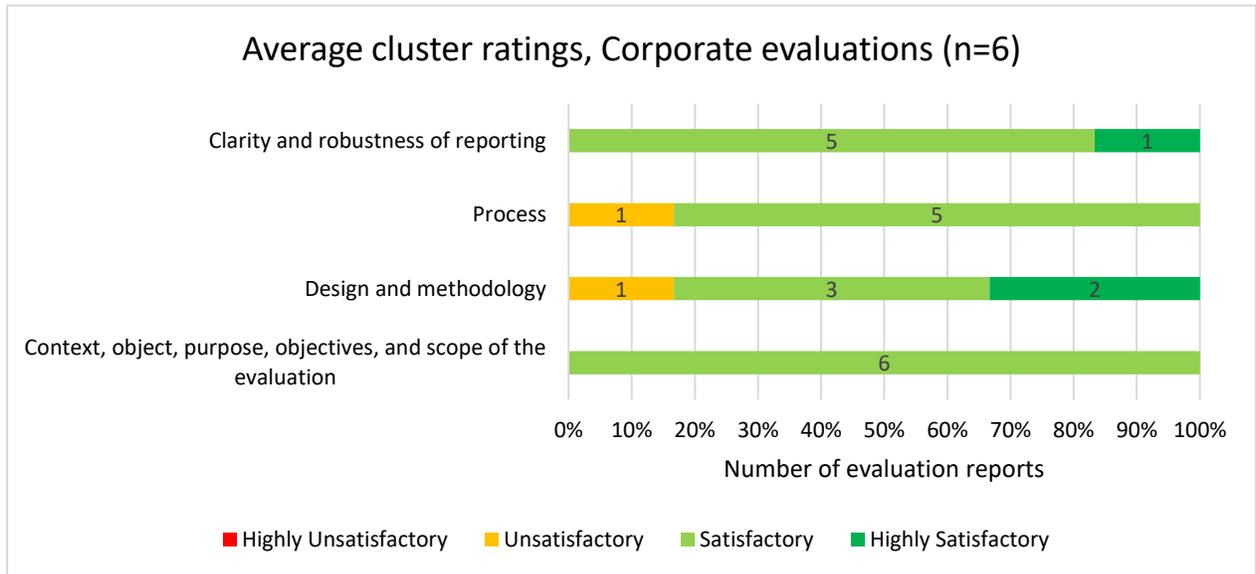
Appendix XI Quality Assessment Results by clusters of criteria, All Evaluations



Appendix XII Quality Assessment Results by cluster of criteria, Country Evaluations

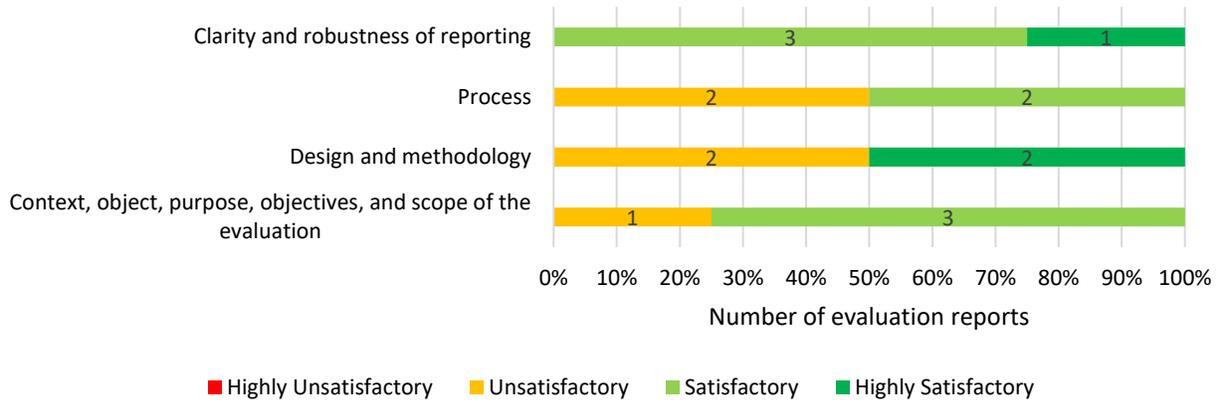


Appendix XIII Quality Assessment Results by cluster of criteria, Corporate Evaluations

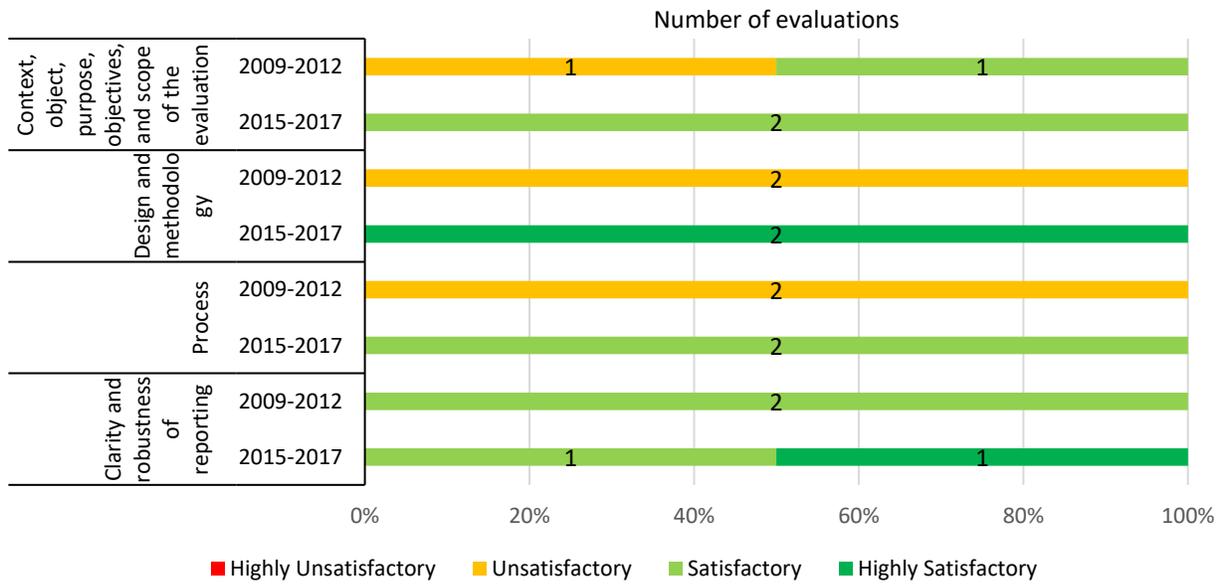


Appendix XIV Quality Assessment Results by cluster of criteria, Thematic/Sector Evaluations

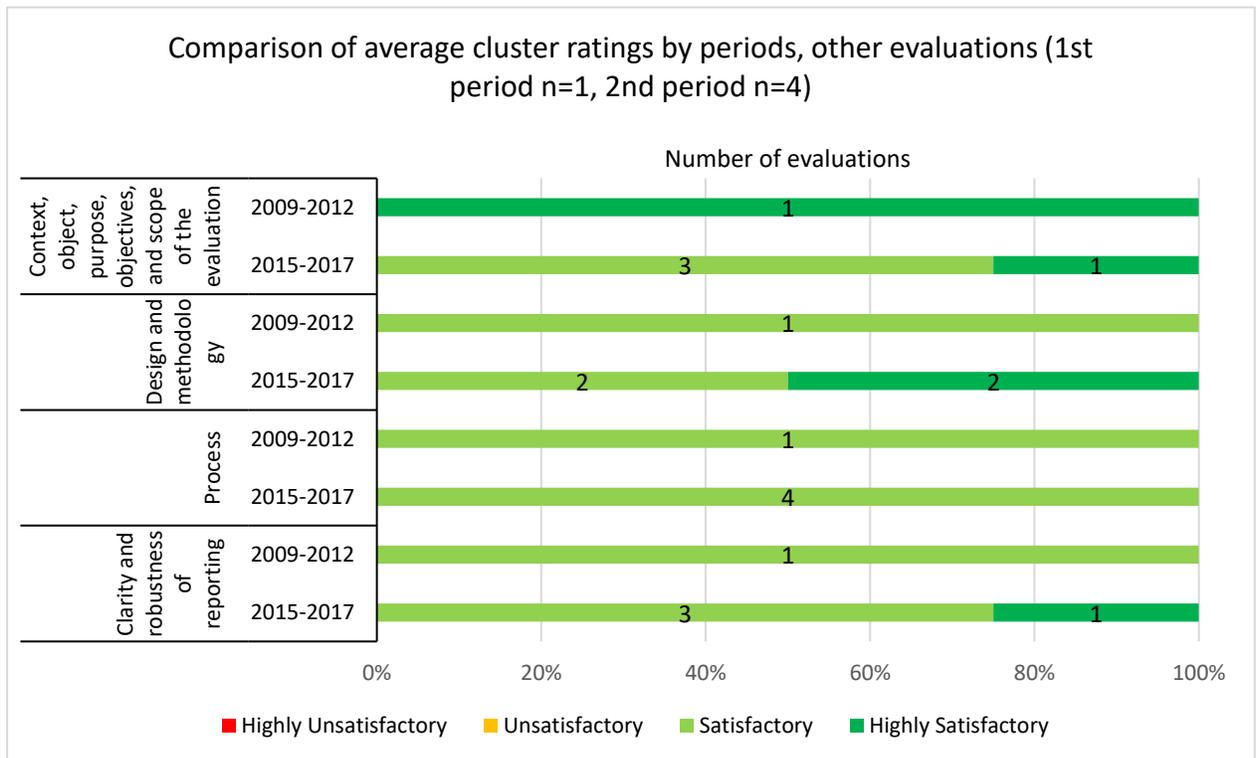
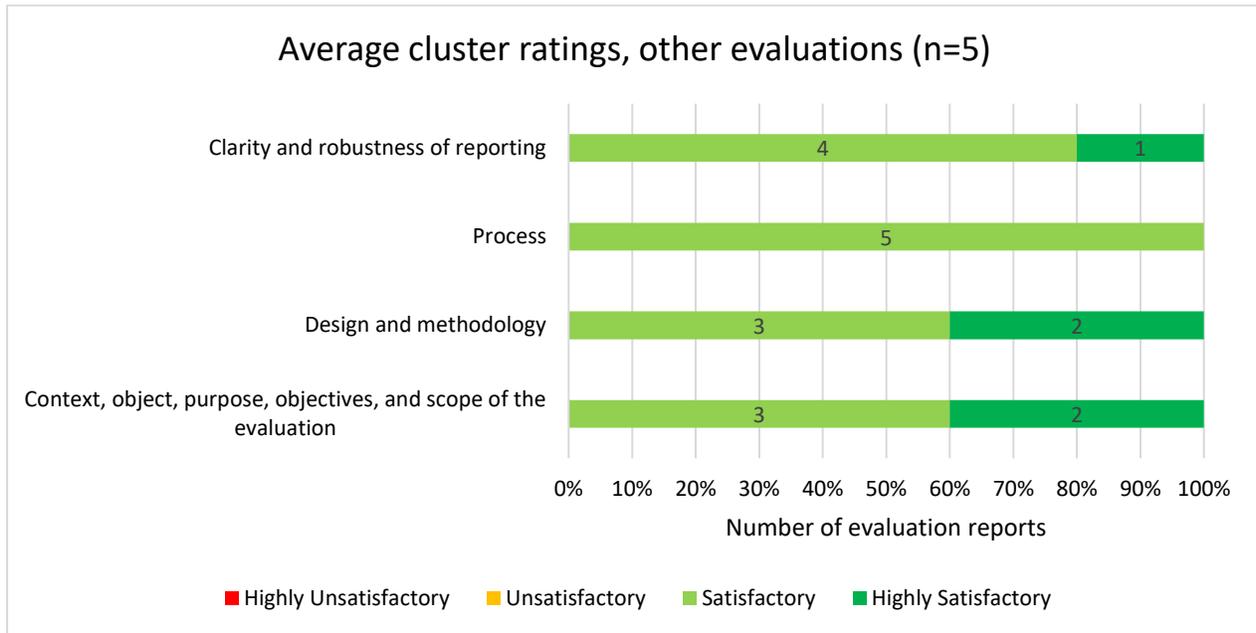
Average cluster ratings, Thematic/Sector evaluations (n=4)



Comparison of average cluster ratings by period, Thematic/Sector evaluations (1st period n=2, 2nd period n=2)

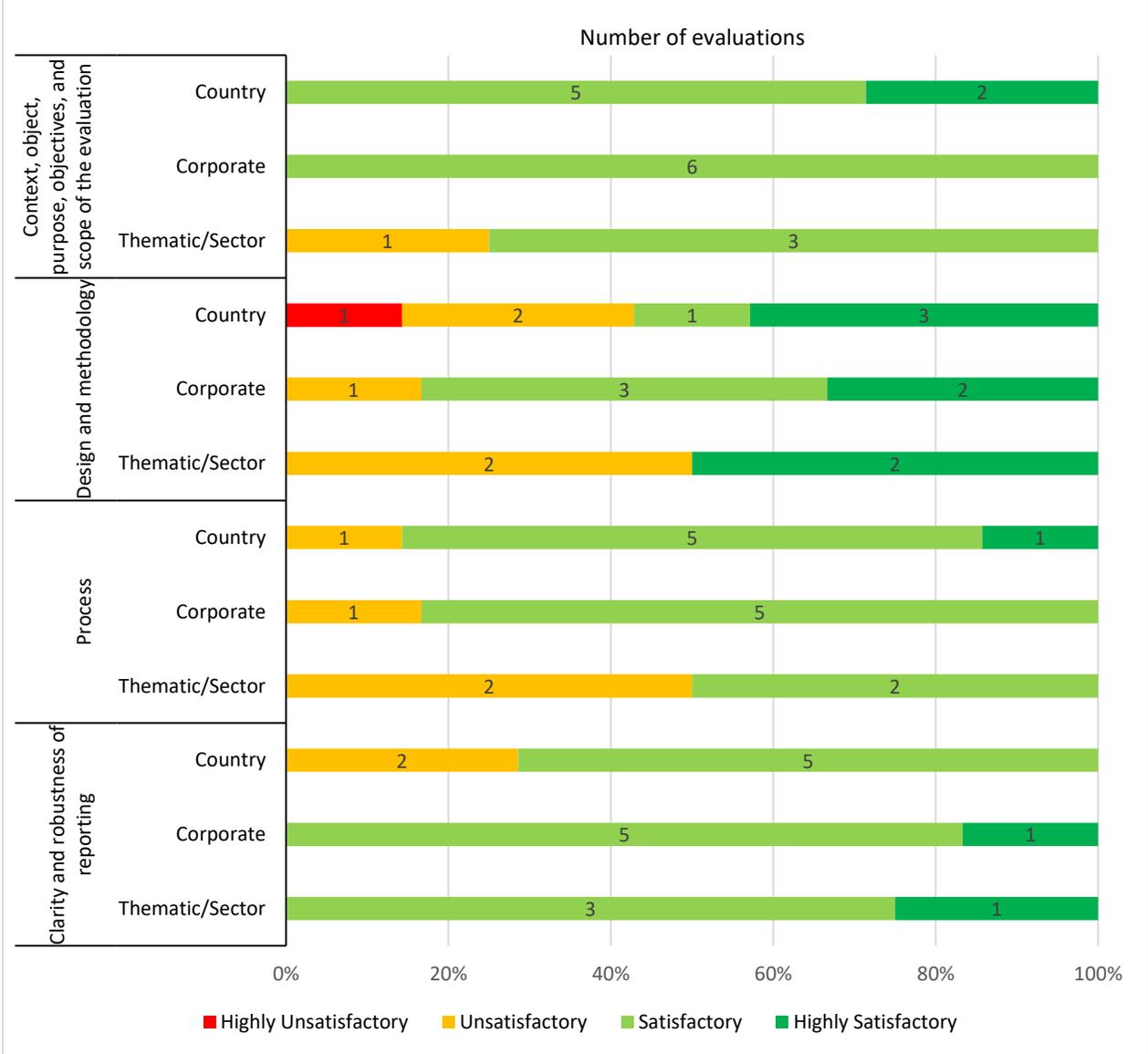


Appendix XV Quality Assessment Results by cluster of criteria, other Evaluations



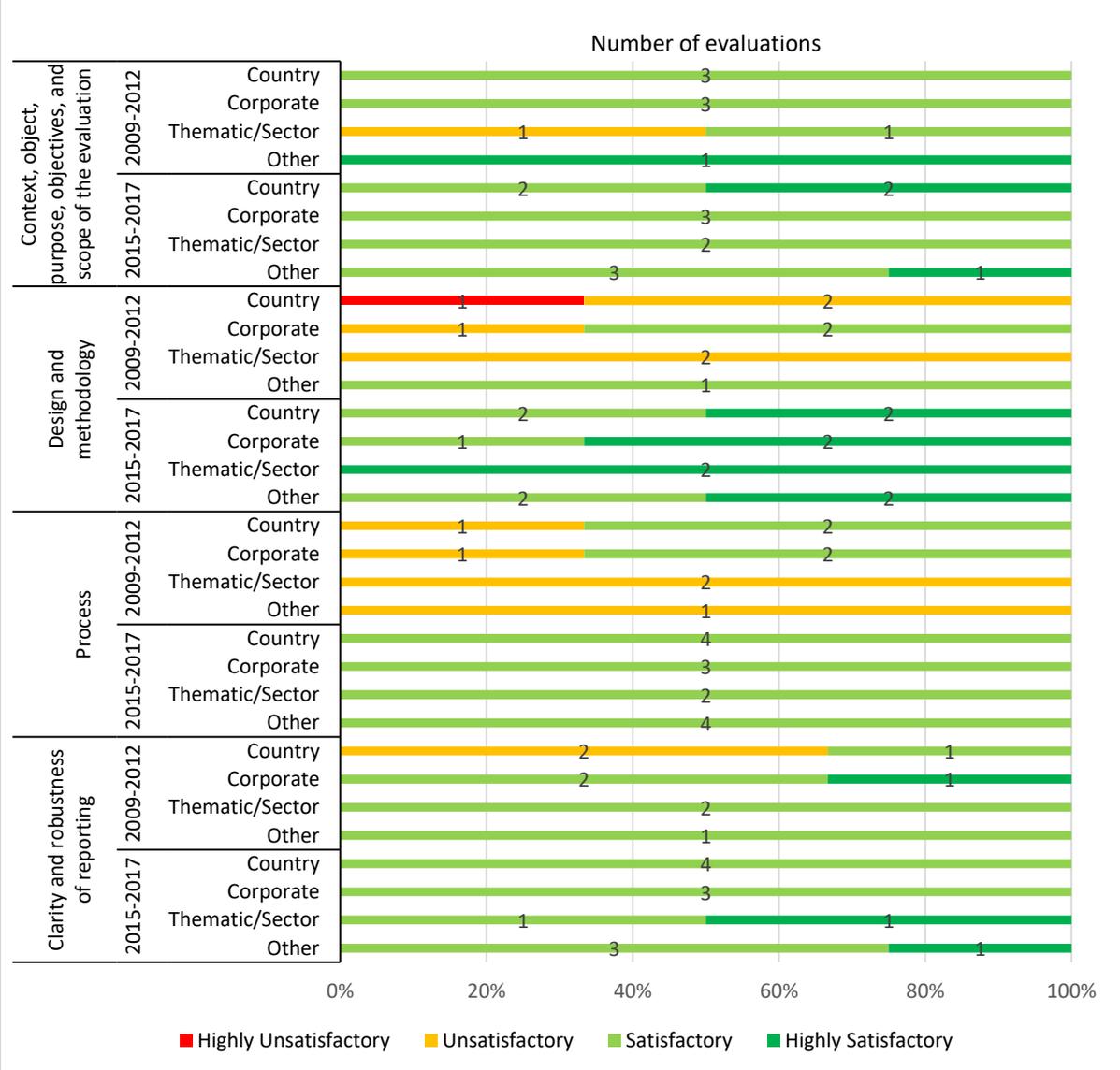
Appendix XVI Quality Assessment Results by cluster of criteria and type of evaluation

Comparison of average cluster ratings by type of evaluation (Country n=7, Corporate n=6, Thematic/Sector n=4))



Appendix XVII Quality Assessment Results by cluster of criteria, evaluation periods, and types of evaluation

Comparison of average cluster ratings by type of evaluation and period (Country n=7, Corporate n=6, Thematic/Sector n=4, Others n=5)



Appendix XVIII List of Documents Consulted

International Standards

Busan Partnership for Effective Development Co-operation: Fourth High Level Forum on Aid Effectiveness, Busan, Republic of Korea, 29 November – 1 December 2011.

Evaluation Cooperation Group. 2017. *Integrating Gender into Project-level Evaluation*. Abidjan: IDEV-AfDB

Evaluation Cooperation Group. 2012. *Big Book on Evaluation Good Practice Standards*. Manila

Evaluation Cooperation Group. 2008. *Good Practice Standards – Country Strategy and Program Evaluations*. Manila

Corporate

African Development Bank Group. 2017. *Annual Development Effectiveness Review 2017 Transforming Africa – Unlocking agriculture’s potential*. Abidjan

African Development Bank. 2013. *At the Center of Africa’s Transformation: Strategy for 2013-2022*. Abidjan

African Development Bank-African Development Fund. June 2001. *The Gender Policy*. OESU.

Independent Development Evaluation. 2013. *African Development Bank Independent Evaluation Strategy 2013–2017*. African Development Bank Group. Abidjan

Independent Development Evaluation. 2013. *Assessing implementation of the AfDB Independent Evaluation Strategy 2013-17 Inception report V3*. African Development Bank Group. Abidjan

Independent Development Evaluation. 2016. *Comprehensive Evaluation of the Development Results of the African Development Bank Group 2004-2013*. Synthesis Report. African Development Bank Group. Abidjan

Country Evaluations

CAMEROUN

- Évaluation de l’Assistance de la Banque, Banque africaine de développement, Fonds africain de développement, novembre 2009.
- Country Assistance Evaluation, Approach Paper, Operations Evaluation Department, 2009.

CAP VERT

- Évaluation de l'Assistance de la Banque 1996-2007, Groupe de la Banque africaine de développement, Département de l'Évaluation des Opérations, octobre 2009.
- Note d'Orientation, Évaluation de l'Assistance de la Banque au Cap Vert, 2009.
- Country Assistance Evaluation, Operations Evaluation Department, October 2009

CÔTE D'IVOIRE

- Management Response, African Development Bank, January 2018.
- Evaluation of the Bank's Country Strategy and Program 2006-2016, Communication and Dissemination Plan, 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement BAD en Côte d'Ivoire 2006-2016, Rapport de synthèse, IDEV, septembre 2017.
- Evaluation of the Bank's Country Strategy and Program 2006-2016, Summary Report, IDEV, September 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement BAD en Côte d'Ivoire 2006-2016, Rapport technique révisé, ADE, juin 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement BAD en Côte d'Ivoire 2006-2016, Rapport technique Volume II - Annexes, ADE, avril 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement BAD en Côte d'Ivoire 2006-2016, Rapport de démarrage, Volume II – Annexes, janvier 2017.
- Projet de pont à péage Henri Konan Bédié, Project Results Assessment, 2017.
- Programme d'Appui à l'amélioration de l'employabilité et de l'insertion des jeunes (PAAEIJ), Project Results Assessment, 2017.
- Projet d'Appui Institutionnel Multisectoriel à la Sortie de Crise (PAIMSC) – Volet Agriculture, Project Results Assessment, 2017.
- Projet d'Appui Institutionnel Multisectoriel à la Sortie de Crise (PAIMSC)- Volet éducation, Project Results Assessment, 2017.
- Projet d'Appui Institutionnel Multisectoriel à la Sortie de Crise (PAIMSC) Volet Santé et VBG, Project Results Assessment, 2017.
- Programme d'appui au renforcement de l'inclusion et de la cohésion sociales (PARICS), Project Results Assessment, 2017.
- Programme de riposte contre Ebola (PRCE), Project Results Assessment, 2017.
- Projet de gestion intégrée du bassin versant du Gourou - phase d'urgence, Project Results Assessment, 2017.
- Programme d'Urgence pour la Restauration des Services Sociaux et Administratifs de Base (PURSSAB), Project Results Assessment, 2017.

- Évaluation des stratégies et programmes de la Banque Africaine de Développement BAD en Côte d'Ivoire, Termes de référence des services de consultance, juillet 2016.

DEMOCRATIC REPUBLIC OF CONGO

- Evaluation of the Bank's Country Strategy and Program 2004-2015, Summary Report, IDEV, March 2017.
- Évaluation de la stratégie et du programme de la Banque 2004-2015, Rapport de Synthèse, IDEV, Mars 2017.
- Management Response, African Development Bank, March 2017.
- Évaluation des stratégies et programmes de la Banque africaine de développement BAD en République démocratique du Congo, 2004-2013, Rapport technique, ADE, juin 2016.
- Évaluation des stratégies et programmes de la Banque africaine de développement BAD en République démocratique du Congo, 2004-2013, Rapport de démarrage – version finale, ADE, octobre 2015.

GABON

- Management Response to the Evaluation of Assistance to Gabon, African Development Bank, April 2011.
- Country Assistance Evaluation, 1996-2008, Operations Evaluation Department, African Development Bank Group, 2011.
- Document d'orientation, Évaluation de l'Assistance de la Banque au Gabon, 2011.
- Évaluation de l'Assistance de la Banque au Gabon, Termes de référence du consultant, 2011.
- Évaluation de l'Assistance de la Banque, 1996-2008, Rapport provisoire, Groupe de la Banque africaine de développement, Département de l'Évaluation des Opérations, août 2009.
- Évaluation de Performance du Multisecteur, 1996-2008, Rapport provisoire, Banque africaine de développement, Département d'Évaluation des Opérations, mars 2009.
- Évaluation de l'Assistance de la Banque au Gabon, Termes de référence de la mission, Groupe de la Banque africaine de développement, décembre 2008.

GHANA

- Evaluation of the Bank's Country Strategy and Program 2002-2015, Summary Report, IDEV, March 2017.
- Country Strategy and Programme Evaluation 2002-2014, Technical Report, BKP, revised July 2016.
- Afram Plains District Agricultural Development Project (APDADP), Project Results Assessment, 2016.
- Export Marketing and Quality Awareness Project (EMQAP), Project Results Assessment, 2016.

- Takoradi II Power Expansion Project, Project Results Assessment, 2016.
- Poverty Reduction and Business Environment Support Programme (PRBESP), Project Results Assessment, 2016.
- Second Poverty Reduction Support Loan (PRSL II), Project Results Assessment, 2016.
- Third Poverty Reduction Support Loan (PRSL III), Project Results Assessment, 2016.
- Multinational Project for The Creation of Sustainable Tsetse and Trypanosomiasis Free Areas in East and West Africa – Ghana Component, Project Results Assessment, 2016.
- Country Strategy and Programme Evaluation 2004-2014, Inception Report, BKP, Draft, revised December 2015.

SOUTH AFRICA

- Management Response, African Development Bank, January 2017.
- Country Strategy and Program Evaluation 2004-2015, Summary Report, IDEV, September 2016.
- Country Strategy and Program Evaluation 2004-2015, Preliminary Dissemination Plan, 2016.
- Independent Evaluation of the AfDB's program and strategy in South Africa 2004-2014, Presentation, IDEV, 2016.
- Country Strategy and Program Evaluation 2004-2014, Final Technical Report, ECORYS – IDEV, May 2016.
- Nedbank Limited – Line of Credit, Project Performance Assessment, 2016.
- Nedbank Limited – Tier II Subordinated Notes, Project Performance Assessment, 2016.
- Standard Bank of South Africa, Project Performance Assessment, 2016.
- Standard Bank of South Africa – Line of Credit, Project Performance Assessment, 2016.
- Industrial Development Corporation Non-Sovereign Guarantee Line of Credit I, Project Performance Assessment, 2016.
- Industrial Development Corporation Non-Sovereign Guarantee Line of Credit II, Project Performance Assessment, 2016.
- Land and Agricultural Bank LoC, Project Performance Assessment, 2016.
- Development Bank of Southern Africa, LoC IV, 2016.
- Transnet Limited, Project Performance Assessment, 2016.
- Medupi Power Station, Project Performance Assessment, 2016.
- Sere Wind Farm, Project Performance Assessment, 2016.
- Country Strategy and Program Evaluation 2004-2014, Inception report – Final, ECORYS – IDEV, October 2015.
- Country Strategy and Program Evaluation 2004-2014, Request for Proposal, African Development Bank, 2015.

UGANDA

- Country Assistance Evaluation, 2001-2007, IEG/OPEV, March 2009.

Corporate Evaluations

Independent Evaluation of the Decentralization Strategy and Process at the African Development Bank, 2009

- Summary Report for ADF-11 Mid Term Review, Background Paper, OPEV, African Development Fund, October 2009.
- Management Response, AfDB/AfDF, July 2009.
- Board Report – Final Draft, IOD/PARC, June 2009.
- Report: First Draft, May 2009.
- Decentralization Evaluation, Most Significant Change: All Stories, OPEV, 2009.
- An Interim Report on The use of Most Significant Change stories, IOD/PARC, January 2009.
- Concept Note, OPEV, Revised December 2008.
- Inception Report, December 2008.
- Terms of Reference for the Evaluation, 2008.

Independent Review of the Quality at Entry of ADF 2005-2008 Operations and Strategies, 2010

- Final Summary Report, AfDBG/OPEV, March 2010
- Management Response, AfDB/AfDF, March 2010.
- Final Technical Report, ITAD, September 2009.
- Revised Inception Report, May 2009.
- Terms of Reference for Consultancy Services, Annex II, 2009.
- Approach Paper.

Project Supervision at the AfDB 2001-2008, 2010

- Evaluation Brief, AfDB/OPEV, July 2010.
- Technical Report, AfDB/AfDF, June 2010.
- Management Response, AfDB/AfDF, March 2010.

- Final Report, AfDB/OPEV, February 2010.
- Technical Annex I, Approach and Methodology, June 2009.
- Technical Annex III, Excerpts from Bank Reports and Manuals on Project Supervision, June 2009.
- Approach Paper, April 2005.

Independent Evaluation of Policy and Strategy Making and Implementation, 2015

- Revised Management Response, AfDB/AfDF, February 2016.
- Summary Report, 2016.
- Commitments Evaluations, Communications and Dissemination Plan, 2015.
- Volume II, Case Studies, Centennial Group, April 2015.
- Inception Report, September 2014.
- Approach Paper, March 2014.

Comprehensive Evaluation of the Development Results of the African Development Bank Group 2004-2013, 2016

- Synthesis Report, IDEV, October 2016.
- Internal Background Paper: A qualitative comparative analysis of the bank's theory of change for the CEDR: Evaluating factors thought to contribute to AfDB's performance at country level, IDEV, May 2016.
- Final Report, IDEV, April 2016.
- Annex B: Evaluation Matrix, 2016.
- Synthesis Methodology Note, IDEV, March 2016.
- Theory of Change, February 2016.
- CEDR Synthesis, Country template, Guidance Note, IDEV, September 2015.
- CEDR Synthesis, Country template, IDEV, September 2015.
- Project Performance Assessment Template, 2015.
- Approach Paper, OPEV, March 2014.
- Exploratory Note, OPEV, October 2013.

The African Development Bank's Human Resource Management Policy and Strategic Directions: A Formative Evaluation, 2017

- Summary Report, Redacted version, IDEV, November 2017.
- Communications and Dissemination Plan, 2017.
- Technical Report, Final (V3), EY, June 2017.
- Emerging Conclusions, Presentation, EY, 2017.
- Inception Report, Version 4, EY, 2017.
- Terms of Reference, 2017.

Thematic/sector Evaluations

Evaluation of Policy-based Operations in the African Development Bank, 1999-2009, 2010

- Management Response, AfDB/AfDF, April 2011.
- Final Report, OPEV/AfDBG, 2011.
- Country Case-Study: Ethiopia, High Level Evaluations Division (OPEV.2), March 2011.
- Country Case-Study: Morocco, High Level Evaluations Division (OPEV.2), March 2011.
- Country Case-Study: Rwanda, High Level Evaluations Division (OPEV.2), March 2011.
- Country Case-Study: Sierra Leone, High Level Evaluations Division (OPEV.2), March 2011.
- Country Case-Study: Tanzania, High Level Evaluations Division (OPEV.2), March 2011.
- Institutional and Policy Review, High Level Evaluations Division (OPEV.2), December 2010.
- Review of Literature and Experience, High Level Evaluations Division (OPEV.2), November 2010.
- Case Study: Botswana Economic Diversification Support Loan, High Level Evaluations Division (OPEV.2), September 2010.
- Case Study: Egypt Financial Sector Reform Programme, High Level Evaluations Division (OPEV.2), September 2010.
- Final Inception Report, OPM/Mokoro/CSEA, May 2010.
- Terms of Reference, 2009.

Unlocking the Potential of Africa's Sub-Regions: Review of Bank Group Assistance to Sub-Regional Development Banks, 2010

- Final Report, Project and Programme Division (OPEV.1), July 2010.

- Inception Report, OPEV, January 2009.

Evaluation of Bank Assistance to Small and Medium Enterprises (2006–2013), 2015

- Management Response, AfDB/AfDF, September 2015.
- Summary Report, Redacted version, IDEV, September 2015.
- Evaluation Technical Report, IDEV, December 2014.
- Draft Inception Report, October 2013.

The African Development Bank's Support for Agricultural Value Chains Development: Lessons for the Feed Africa Strategy, 2017

- Communications and Dissemination Plan, 2017.
- Summary Report, IDEV, December 2017.
- Final Technical Report, SDF Global, October 2017.
- Case studies: Kenya report (Final), ADE, August 2017.
- Case studies: Liberia report (Final), ADE, August 2017.
- Case studies: Morocco report (Final), ADE, August 2017.
- Case studies: Mozambique report (Final), ADE, August 2017.
- Case studies: Rwanda report (Final), ADE, August 2017.
- Case studies: Uganda report (Final), ADE, August 2017.
- Case studies: Zambia report (Final), ADE, August 2017.
- Case studies: Ivory Coast report (Final), ADE, August 2017.
- Draft Inception Report, IDEV, December 2016.
- Approach Paper, 2016.

Others: Synthesis

Mainstreaming Gender Equality: A Road to Results or a Road to Nowhere? 2012

- Evaluation Synthesis, Presentation, 2012.
- Synthesis Report, OPEV, 2012.

- Approach Paper, OPEV, 2010.
- Inter-Office Memorandum, AfDBG, June 2010.

Towards Private Sector Led Growth: Lessons of Experience - Evaluation Synthesis Report, 2016

- Evaluation Synthesis Report, IDEV, October 2016.
- Evaluation Synthesis Report Draft, IDEV, October 2016.
- Evaluation Synthesis Report, Revised Draft, Centennial Group, July 2016.
- Inception Report, Centennial Group, April 2016.
- Terms of Reference, 2016.

Others: Impact

Impact Evaluation of the Rural Water Supply and Sanitation Program in Ethiopia 2006-2014, 2016

- Management Response, AfDB/AfDF, October 2016.
- Summary Report, IDEV, September 2016.
- Report, IDEV, September 2016.
- Inception Report, October 2014.
- Concept note OPEV impact evaluations AfDB Group assistance: 1. Ethiopia Rural Water Supply and Sanitation; 2. Tanzania Energy sector, OPEV, March 2014.
- Draft Approach Paper, 2014.

Others: Regional

Evaluation of the Regional Integration Strategy and Operations of the African Development Bank, 2011-2016, 2017

- Plan de dissémination des résultats, 2017.
- Synthesis Report, IDEV, December 2017.
- Case study: Evaluation of the Congo Basin Ecosystems Conservation Support Programme (PACEBCo), Final version, November 2017.

- Case study: Evaluation of the Institutional Capacity Building Support to the Economic Community of Central African States (PARCI-ECCAS), Final version, November 2017.
- Case study: Evaluation of the Lake Chad Basin Sustainable Development Programme (PRODEBALT), Final version, November 2017.
- Analyse du portefeuille des opérations multinationales, Novembre 2017.
- Case study: Evaluation of Ketta-Djoug Road Project, Final version, October 2017.
- Review of the Regional Integration Strategy Paper (RISP), Final version, September 2017.
- Quality at Entry Review, Final version, July 2017.
- Challenges and Opportunities for Regional Integration in Central Africa, Final, December 2016.
- Concepts of Regionalism and their Use in the Bank's Policy and Strategy Documents, December 2016.
- Document d'Orientation, Décembre 2016.

Others: Cluster

Spurring Local Socio-economic Development through Rural Electrification, 2017

- Communications and Dissemination Plan, 2017.
- Report, IDEV, November 2017.
- Synthesis Report, IDEV February 2016.
- Common Data Collection Protocol for Project Performance Evaluation, February 2015.
- Cluster Evaluation Approach Paper, Draft, January 2015.

Appendix XIX Abidjan Workshop Agenda and Report



Workshop Presentation: Quality Assessment of a Sample of Evaluation Products IDEV-AfDB

Agenda March 27, 2018

-
- | | |
|----------------------------|--|
| 9.0h-9.10h | Workshop Introduction: Mr. Rakesh Nangia, Evaluator General, IDEV |
| 9.10-9.20h | Introductions (All Participants) |
| 9.20-9.30h | Review of Objectives and Steps of Quality Assessment (Penny and Anne-Marie) |
| 9.30-9.40h | Overview of Workshop Objectives and Agenda (Anne-Marie) |
| 9.40-10.30h | <p>Interactive PowerPoint Presentation: Main Findings and Recommendations of Quality Assessment (Anne-Marie with questions from participants)</p> <p>a) Strengths and weaknesses of evaluations 2009-2013 vs. 2015-2017 by theme:</p> <ul style="list-style-type: none"> • Context, object, purpose, objectives, and scope of the evaluation • Design and methodology • Process • Clarity and robustness of reporting <p>b) Any issues need clarification?</p> |
| <i>Break (10.30-10.45)</i> | |
| 10.45-12.30h | <p>Discussion on Quality Assessment Recommendations, including Implications for Knowledge Management and Communications (Five Thematic Sub-groups)</p> <p>a) Do QA recommendations related to each of the four key themes identified resonate with participants?</p> <p>b) Moving forward - How can recommendations:</p> <ul style="list-style-type: none"> • Be implemented and support learning? • Inform the next five-year strategy? • Help IDEV identify low hanging fruit versus long term investments? |

Lunch (12.15-14.00)

- 14.0-15.0h Report back from the five groups and discussion with full team
- 15h-15.30h Questions for Participants in View of Finalizing the Synthesis Report (Anne-Marie)
- 15.30-16.0h Discussion on next steps for IDEV to make use of these findings in practice
- 16.0-16.15 Workshop Wrap-Up and Evaluation

Summary

The workshop with IDEV personnel held at AfDB headquarters in Abidjan on March 27, 2018 was essentially an interactive knowledge-sharing workshop to discuss the process, findings, and recommendations of the quality assessment conducted by Universalialia. Ms. Anne-Marie Dawson represented the assessment team at the workshop which was co-facilitated by Ms. Penelope Jackson, Chief Evaluation Officer at IDEV. Following an introduction to the workshop by Mr. Rakesh Nangia, IDEV Evaluator General, and introductions of some 25 IDEV personnel attending, the QA process, methodology, main findings, and recommendations were presented in a PowerPoint and involved discussion with participants on a variety of issues. These included the QA methodology, data analysis, institutional mechanisms for communicating evaluation findings and for following up on recommendations among other things.

Following the PowerPoint presentation and discussion, the participants broke into five sub-groups, each of which was tasked with examining one specific QA recommendation, exchanging views on the issue it addressed, and exploring the ways it could be practically implemented. Following the lunch break, a rapporteur from each sub-group presented the results of his/her group's discussion to the full team. This allowed all participants to learn about each other's existing practices, e.g. strategies for building the capacities of local evaluators in the context of various types of evaluations, as well as to discuss ways existing good practices could be further embedded in processes and recommendations for improving practices could be concretely applied. On the whole, the participants agreed with the recommendations and shared insights for refining them to maximize their utility. The workshop evaluation questionnaires anonymously completed by almost all participants highlight their appreciation of the opportunity to provide feedback on the QA recommendations and to meet as a group to share ideas about their work and learn from one another.

Following is the detailed workshop report produced by IDEV, including the list of participants.

**Workshop on sharing the findings of the Quality Assessment of a sample of IDEV evaluations,
March 27, 2018**

Team members: Anne-Marie Dawson (Universalialia), Penelope Jackson & Latefa Camara (IDEV)

Workshop report**1. Objective of the workshop**

The main objective of this workshop was to share with the IDEV team the results from the quality assessment exercise of a sample of IDEV's evaluation products. This was conducted by

an external firm (Universalialia) of consultants for credibility and impartiality reasons. The QA assessment is part of the broader assessment of IDEV's implementation of the 5-year Independent Evaluation Strategy, 2013-2017. The workshop therefore sought to get the views of IDEV colleagues on what emerged from the QA including draft recommendations

2. Introduction and presentation

Following the introductory remarks by the Evaluator General, and the introduction of each participant, Anne-Marie, team leader for the Universalialia team presented the main findings of the quality assessment exercise. A few suggestions were made by the participants regarding the QA, such as replacing the terms baseline and endline with other terms less statistical like first period, second period. The fact that the sample size does not allow for statistical tests was also noted.

The presentation showed that, overall, Universalialia found that there has been improvement in the quality of the evaluations that IDEV has produced over the years, principally comparing evaluations in the first period (2009-2012) and second period (2015-2017).

One participant highlighted the fact that there have been many other changes in IDEV –apart from the strategy -that may have contributed to this quality improvement including increase in budget for evaluations in recent years. It was clarified that attributing any changes to the strategy or other was not within the brief of the assignment given to Universalialia. Participants also pointed to institutional restrictions –such as the number of pages of reports presented to the Board might limit the amount of information included in “summary reports” as opposed to other background and design documentation.

3. Outcome of Group discussions

Following the presentation and the Q&A, the workshop participants were divided into different groups for more in-depth discussion based on the different recommendations made by the Universalialia team. The discussions in the group work were interesting thanks to the active engagement of the participants. Of the 8 recommendations made, the groups had the opportunity to discuss five, based on what participants judged more relevant and important for IDEV's work. The following are the main outcomes or issues discussed during this group work:

- ⇒ **Recommendation:** *Consider adding a requirement for stakeholder mapping. This requirement would be stipulated in the ToR and the stakeholder map developed by the evaluators and presented in both the inception and evaluation reports (in annex if too detailed). This would ensure all key stakeholders are clearly identified and their roles and contributions clearly described.*

Group conclusion: The team agreed overall with this recommendation, and believe this will reinforce transparency and the use of evaluation. The approach to stakeholder mapping could

be set out in the planned IDEV evaluation manual at approach paper/inception report. It is important to identify the stakeholders based on the potential sources of information for the evaluation, as well as their different information needs

⇒ **Recommendation:** *Ensure effective cross-cutting integration of issues related to gender equality, inclusive growth, environment/climate change, fragility that are explicitly identified as areas of focus in AfDB's ten-year strategy. The findings reveal uneven coverage of these dimensions both within each type of evaluation product and across the range of products. These dimensions are commonly referred to as "cross-cutting" yet they are often not always effectively integrated in a cross-cutting manner in evaluations.*

Group conclusion: IDEV should make sure it identifies the cross-cutting issues that matter for the department and ensure that for each evaluation, the relevant cross-cutting issue is identified and considered. It is important to note that not all cross-cutting issues are relevant for every type of evaluation we do. The integration of cross-cutting issues depends also on the design of operations, sometimes there are no data even if the evaluation team wants to do something on these issues, as the design or implementation/monitoring did not take this into account. There may, however, be opportunities to assess some cross cutting issues in any case and we should examine this on a case by case basis.

⇒ **Recommendation:** *Ensure that OECD-DAC and AfDB-IDEV evaluation principles related to evaluation independence, impartiality, ethics, partnership, coordination, and capacity building are clearly and explicitly addressed in the various evaluation products and considered as part and parcel of the evaluative process. The AP/IR should include a standard section with respect to the consideration of these principles.*

Group conclusion: This was considered an important recommendation by the group, though the difference between what is recorded and what is taken for granted was noted (e.g. independence and impartiality already being spelt out in the Policy).

IDEV evaluations should for example ensure that capacity building is a component, by engaging local consultants in evaluations. This is an interesting point but some questions were raised with regards to this. For example, in practice implementing such initiative to build capacity can be very time consuming during an evaluation. The best way to do this is to plan it well and include in the design of the evaluation so that there is sufficient time allocated to the entire evaluation to allow capacity building to happen. It is a very heavy process both administratively and in terms of time also. However, sometimes, the capacity building that IDEV does for example at the level of ministries is a different thing. It is not always easy to involve these people in our evaluation due to the independent nature of what we do.

IDEV should apply some practices in research to its evaluation. For example the ethical issues should be included and mentioned during our survey or focus groups/interviews.

- ⇒ **Recommendation:** *To limit overlap between successive products for each evaluation, consider replacing approach papers with evaluation ToR that are less prescriptive and requiring that inception reports be systematically produced by external consultants detailing the evaluation methodology and workplan.*

Group conclusion:

It is important that IDEV makes the distinction about the different types of documents that IDEV produces in the course of an evaluation, and make clear what each document should contain. There is a need for an IDEV handbook that explains every product we deliver and in what circumstances they are required. The production of inception report and approach paper versus only inception report for example depends on each evaluation. For example, with a detailed enough approach paper, a team may feel unnecessary to produce an inception report in addition; or, where using a range of individual consultants, may ask those consultants to produce inception reports for their specific components. Some feel that the approach paper should be done by the IDEV team, and the inception report will be only a completion of what is available by the external consultants that should add more value. Some felt there is more scope to complete inception reports internally, some felt there was a risk where using a firm which had not bought into the methodology through involvement in the inception phase.

These different views reflect the fact that IDEV has more than one model for how it involves consultants. It is possible to contract out all the data collection and it is possible to contract out only specific components; it is possible to use firms or individuals. What documents are produced needs to relate to what works for each model. This needs to be clearly documented.

- ⇒ **Recommendation:** *Ensure a clear distinction is made between evaluation findings, conclusions, and lessons learned. The findings with respect to the evaluation reports reviewed sometimes indicate overlap between these elements, i.e. findings including evaluator judgments and performance ratings, conclusions essentially repeating key findings, and lessons learned worded like conclusions, even recommendations*

Group conclusion: It is important that the evaluators in IDEV have a clear definition of what each concept means. This will ensure that there is a clear and logical flow from the findings to the conclusions and to recommendations. The recommendations should be tied to the conclusions. Lessons are important, but should all four aspects (Findings, conclusions, lessons and recommendations) be in an evaluation?

4. Way forward

Based on the various discussions, the team at Universalia will review some aspects of the QA report before submission of a new draft to IDEV. Issues discussed today will be taken into account both in the assessments and in the recommendations. This draft will be shared with the full team. The assessment of individual evaluations will be shared with individual task managers.

The final version of the quality assessment report will feed into the overall self-assessment of the implementation of the IDEV strategy.

Annex 1: Overview of workshop participant feedback

Overall, the workshop was well received by participants. More than 80% of feedback agreed or strongly agreed that the workshop objectives were clear and relevant. They also agreed on the fact that sufficient time and space was given to express views.

The most interesting parts of the workshop were found to be:

- The group discussions were found to be the most interesting aspects of the workshop, allowing staff to express views on specific themes of interest. Discussions on the recommendations were found to be particularly interesting by participants. The opportunity to discuss recommendations from the quality assessment was very much appreciated.

Main takeaways from the workshop

The most interesting takeaways were:

- The importance of the strategy in improving outcomes
- The importance of team work
- The integration of cross-cutting themes in evaluations
- Discussions on challenges faced by evaluators and
- Reminder of the need to maintain rigor in our evaluations

Annex 2: List of participants

N.	Name	Position
1	Akua Arthur-Kissi	Evaluation Officer-IDEV2
2	Anne-Marie Dawson	Universalialia Team leader
3	Brenda Tsausagae	Consultant-IDEV1
4	Carla Felix	Consultant –IDEV2
5	Clement Banse	Evaluation Officer, IDEV 1
6	Daniel Andoh	Consultant-IDEV3
7	Eneas Gakusi	Chief Evaluation Officer-IDEV1

8	Eric Yai	Consultant-IDEV2
9	Eustace Uzor	Consultant-IDEV2
10	Foday Turay	Chief Evaluation Officer, OIC IDEV1
11	Gilbert Adjimoti	Consultant, IDEV 2
12	Girma Kumbi	Principal Evaluation Officer, IDEV2
13	Innocent Bledou	Consultant, IDEV 2
14	Joseph Mouanda	Principal Evaluation Officer-IDEV1
15	Karen Rot-Munstermann	Division Manager IDEV3
16	Latefa Camara	Consultant-IDEV2
17	Mabarakissa Diomande	Evaluation Officer-IDEV1
18	Madhu Mampuzhasseril	Chief Evaluation Officer, OIC IDEV 2
19	Mireille Cobinah	Evaluation Knowledge Assistant, IDEV3
20	Mohamed Coulibaly	Consultant-IDEV1
21	Monica Lomena-Gelis	Principal Evaluation Officer IDEV2
22	Nagnouma Kone	Consultant-IDEV 3
23	Najade Lindsay	Consultant-IDEV3
24	Penelope Jackson	Chief Evaluation Officer-IDEV1
25	Rakesh Nangia	Evaluator General
26	Souleymane Dieye	Senior Database Assistant, IDEV3
27	Stephanie Yoboue	Consultant-IDEV 1
28	Telesphore Some	Consultant-IDEV3