GREEN CLIMATE FUND | Independent Evaluation Unit

2026

EVALUABILITY ASSESSMENT IV

The Fourth Evaluability assessment of the Green Climate Fund's funding proposals

Hanna Baek, Soledad Barone, Martin Dellavedova, Alejandro González-Caro, Eric Müller, Andreas Reumann, Susumu Yoshida

The Fourth Evaluability Assessment of the Green Climate Fund's Funding Proposals

Hanna Baek, Soledad Barone, Martín Dellavedova, Alejandro González-Caro, Eric Müller, Andreas Reumann, Susumu Yoshida

01/2026

175, Art center-daero
Yeonsu-gu, Incheon 22004
Republic of Korea
Tel. (+82) 032-458-6450
Email: ieu@gcfund.org
https://ieu.greenclimate.fund

First Edition

This paper is a product of the Independent Evaluation Unit at the Green Climate Fund (GCF/IEU). It is part of a larger effort to provide open access to its research and work and to make a contribution to climate change discussions around the world.

While the IEU has undertaken every effort to ensure the data in this report is accurate, it is the reader's responsibility to determine if any and all information provided by the IEU is correct and verified. Neither the author(s) of this document nor anyone connected with the IEU or the GCF can be held responsible for how the information herein is used.

Citation
The suggested citation for this paper is:
Baek, Hanna et al. *The Fourth Evaluability Assessment of the Green Climate Fund's Funding Proposals*. Learning paper. Independent Evaluation Unit, Green Climate Fund, January 2026.

Credits
Head of the GCF Independent Evaluation Unit: Andreas Reumann
Evaluation task lead: Susumu Yoshida
Editing: Beverley Mitchell
Layout: Giang Pham
Cover design: Therese Gonzaga

A FREE PUBLICATION

Printed on eco-friendly paper

Acknowledgements

The authors express their sincere gratitude to SISTME for their rigorous and exceptionally detailed analytical work throughout this evaluability assessment. The SISTME team's meticulous approach to analysing and reasoning through each indicator for every project has provided not only the essential foundation for this assessment but also critical insights that will shape future evaluability assessment rounds. Their comprehensive brainstorming sessions and collaborative efforts in systematically examining the four studied evaluability dimensions across the GCF portfolio have been instrumental in developing a nuanced understanding of measurement challenges and opportunities in climate finance.

Our deep appreciation extends to the broader GCF IEU team, whose collective expertise and dedication over the years have continuously strengthened this evaluability study. Their cumulative contributions have built the institutional knowledge and methodological frameworks that made this fourth evaluability assessment possible.

This report represents a collaborative effort to advance the GCF's learning agenda at a critical juncture in climate finance. By mapping the evolving landscape of project evaluability, we aim to contribute to the larger conversation about measuring what matters in climate action, ensuring that each investment not only addresses immediate climate needs but also builds our collective knowledge about effective interventions. The authors are grateful to all who have contributed to this pioneering work in strengthening the evidence base for climate finance decisions. All remaining errors are the sole responsibility of the authors.

List of authors

The authors of this paper are listed in alphabetical order by last name.

FULL NAME	AFFILIATION
Hanna Baek	Independent Evaluation Unit, Green Climate Fund
Soledad Barone	SISTME
Martin Dellavedova	SISTME
Alejandro González-Caro	Independent Evaluation Unit, Green Climate Fund
Eric Müller	SISTME
Andreas Reumann	Independent Evaluation Unit, Green Climate Fund
Susumu Yoshida	Independent Evaluation Unit, Green Climate Fund

Table of Contents

List of Tables

List of Figures

List of Boxes

List of abbreviations

AE	Accredited entity
B.40	The fortieth meeting of the Board
DAE	Direct access entity
EQA	Evaluation quality assessment
FP	Funding proposal
GCF	Green Climate Fund
IAE	International accredited entity
IE	Impact evaluation
IEU	Independent Evaluation Unit
IRM	Initial resource mobilization
M&E	Monitoring and evaluation
PAP	Proposal approval process
pp	Percentage point
SAP	Simplified approval process

I. Introduction

1. The Green Climate Fund (GCF) is a multilateral fund created to make significant and ambitious contributions to global efforts to combat climate change. The GCF contributes to achieving the objectives of the United Nations Framework Convention on Climate Change and the Paris Agreement. It aims to promote a paradigm shift towards low-emission and climate-resilient development pathways by helping developing countries reduce their greenhouse gas emissions while supporting countries' specific needs in adapting to and combating climate change's adverse effects. For developing countries, the GCF provides support through various financial modalities, including grants, loans, and market instruments such as bonds and equity.

2. Following the first assessment conducted in 2019, this document assesses the quality of proposals approved for financing by the GCF (funding proposals (FPs)). It asks the following question: **To what extent are GCF-supported programmes and projects capable of credibly reporting their impacts, effectiveness, and efficiency in an evidence-based and robust way?**

3. We ask this question for two reasons. First, the GCF's overall goal is to support a paradigm shift towards low-carbon, high-resilience pathways. Therefore, it is critical to understand if a paradigm shift is occurring and how much of this shift is attributable to the GCF. The Fund's contribution to this shift depends on its investments credibly committing to, and measuring, the results they explicitly aim to achieve. Second, measurement in the climate change space is difficult. Climate change action requires that large numbers of people act simultaneously to individually effect change that together must represent a large enough and critical change to make a difference. Results from individual actions on overall global climate change will only be apparent after hundreds of years, if not longer. However, the evaluation team can assess the extent to which current investments are likely to yield these results. It is important the GCF examines projects for the likelihood of these results. This is to understand the probability of success and the credibility of results reporting (should it occur) and enable the GCF to reliably report its overall contribution to this climate action effort. It is even more important to assess, test and establish the credibility of these results.

4. The GCF invests its resources using several criteria. Among these are the investment criteria, which require that projects show proof of impact potential, sustainability, paradigm shift potential, country ownership, climate relevance, effectiveness and efficiency. These are also among the criteria the GCF's Independent Evaluation Unit (IEU) uses to assess the quality of the GCF portfolio's performance, activities and results.

5. This study presents the results of an IEU desk assessment of the GCF portfolio to assess the evaluability of its funded activities. The study builds on the findings of three prior assessments, conducted in 2019,[1] 2022[2] and 2024.[3] The study has two main aims. First, to assess the quality of the FPs that the GCF has approved and is currently supporting. The findings are intended to help project developers and managers learn from past experiences and design stronger proposals in the future, proposals with a higher likelihood of producing measurable results and achieving success. Second, the study aims to inform the GCF investment criteria, introduce evidence-based learning opportunities into GCF projects and processes, and inform the implementation and overall impact of GCF resources.

6. These two aims serve three purposes. First, to help inform, where possible, risks that may arise in currently supported projects and to alert project managers. Second, to improve the

[1] Fiala et al., *Becoming Bigger, Better, Smarter: A Summary of the Evaluability of Green Climate Fund Proposals* (2019).
[2] Fiala et al., *Evaluability Assessment of the Green Climate Fund Funding Proposals* (2022).
[3] González-Caro, *The Third Evaluability Assessment of the Green Climate Fund's Funding Proposals* (2025).

overall quality of proposals. Third, to support projects in enhancing their measurement approaches and in exploring methods that FPs may apply for this purpose. It is expected that this discussion and the use of robust methods will enable the GCF to report its overall impact in a measurable and credible manner.

II. Summary of the GCF's funded projects as of December 31, 2024

2.1 Overview of the GCF's portfolio

7. Projects represent the primary mechanism through which the GCF invests in low-emission, high-resilience development pathways. All GCF-supported activities must demonstrate climate rationale to receive funding. The GCF aims to drive paradigm shifts in both climate mitigation and climate adaptation efforts. Approved projects are classified into three categories: mitigation, adaptation, and cross-cutting:

(a) Mitigation projects help developing countries reduce their greenhouse gas emissions.

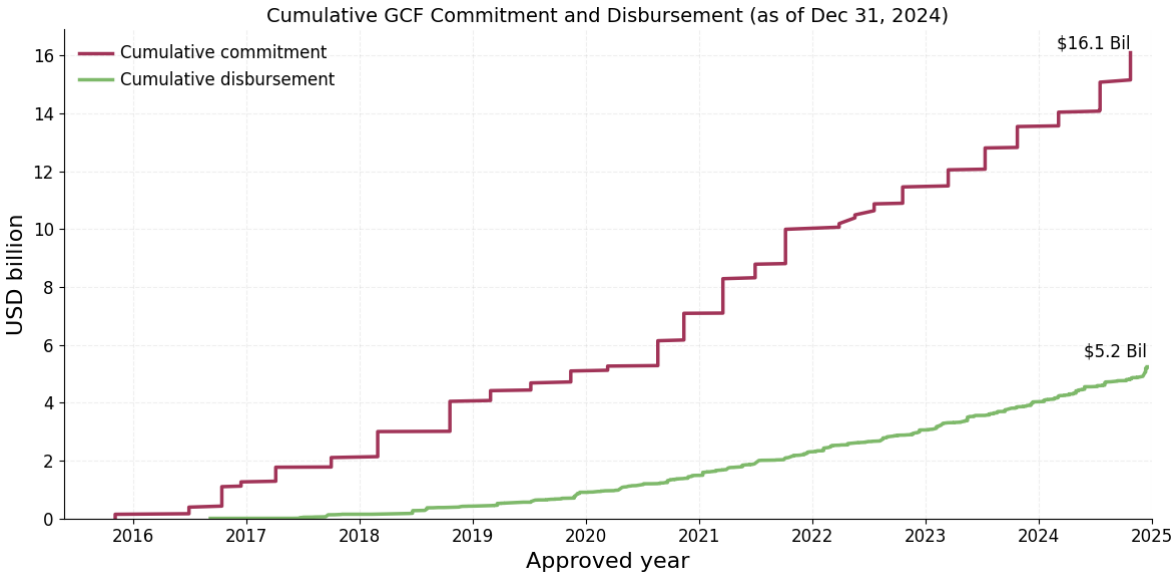
(b) Adaptation projects enhance countries' abilities to withstand climate and weather shocks while increasing community resilience.

(c) Cross-cutting projects simultaneously address both mitigation and adaptation objectives.

8. This analysis examines the GCF's approved project portfolio, consisting of 286 projects (including both FPs and simplified approval process (SAP) projects), approved up to the fortieth meeting of the Board (B.40) in 2024. This total reflects all approved projects maintained in the portfolio excluding: FP031 (never submitted); FP032, FP079, FP088 (now FP110), and FP123 (withdrawn by accredited entities (AEs)); FP029, FP030, FP006, FP038, FP054, FP065, FP104 (approvals lapsed); and FP146 (terminated).

9. Of the 286 projects in the approved portfolio, 249 (approximately 87 per cent) have effective funded activity agreements and have entered the implementation phase, while the remainder were still in post-approval stages as of 31 December 2024. The portfolio has achieved a 33 per cent disbursement rate, with USD 5.2 billion disbursed for funded activities as of 31 December 2024 (see Figure 1).

Figure 1. Cumulative GCF commitment and disbursement



Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including all projects approved through B.40.

2.2 Portfolio by theme – mitigation versus adaptation funding

10. In this analysis, adaptation funding encompasses both stand-alone adaptation projects and the adaptation component of cross-cutting projects. Similarly, mitigation funding includes stand-alone mitigation projects and the mitigation component of cross-cutting projects. This approach provides a comprehensive view of how resources are allocated across climate action themes.

11. Mitigation activities make up 54 per cent of the total funding in the GCF approved portfolio, while adaptation accounts for 46 per cent of the funding, as per the cut-off date of 31 December 2024. This distribution underscores several key trends in climate finance allocation (see Figure 2):

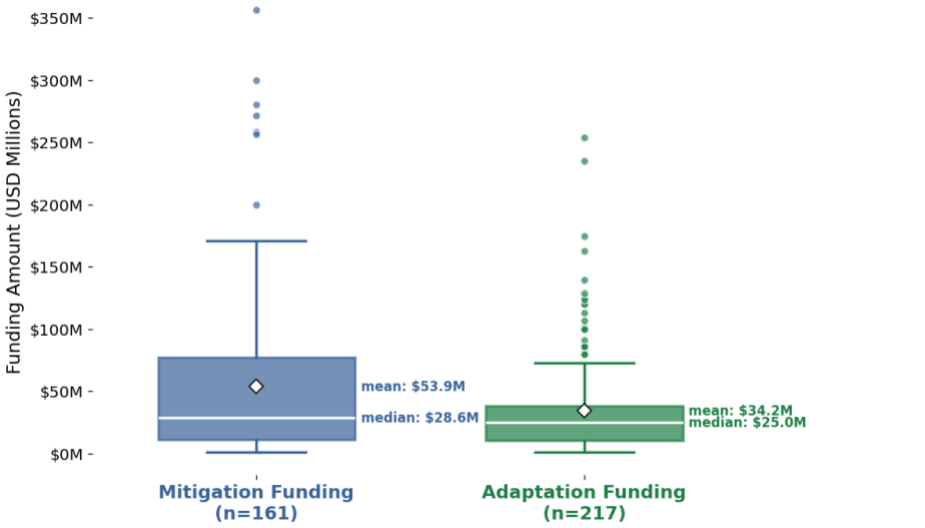
(a) **Project range:** Mitigation funding ranges from USD 0.76 million to USD 356 million, while adaptation funding ranges from USD 0.84 million to USD 253.8 million.

(b) **Average funding:** The average funding for mitigation projects is substantially higher (USD 53.9 million) compared to adaptation projects (USD 34.2 million), indicating strategic emphasis on larger-scale mitigation investments.

(c) **Median values:** The median funding for mitigation projects is USD 28.6 million, while that of adaptation projects is USD 25 million. This relatively moderate difference in medians indicates that the higher average funding for mitigation is driven by some very large projects.

(d) **Total GCF approved portfolio:** The analysis encompasses a total of 286 projects in the approved portfolio, with 161 projects with mitigation components (amounting to USD 8,673.1 million) and 217 projects having adaptation components (totalling USD 7,425.2 million). It is important to note that these figures include 92 cross-cutting projects that address both adaptation and mitigation objectives, 69 projects with only mitigation component, and 125 with only adaptation component, making a total of 286 projects.

Figure 2. GCF funding distribution and comparison of adaptation and mitigation projects



Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including all projects approved through B.40.
Note: Cross-cutting projects are represented in both mitigation and adaptation categories according to their result area investment weights as specified in the GCF Secretariat's public API data. Therefore, the sum of projects across both categories exceeds the total number of approved projects (286), as cross-cutting projects with both components are counted proportionally in each category.

12. The boxplot visualization shows the greater spread and higher outliers in mitigation funding compared to adaptation funding, while also showing the difference in median values. The white diamonds represent the mean values, illustrating how these are pulled upward by large outliers, particularly in the mitigation portfolio.

2.3 Funding by GCF region

13. There is a significant disparity between the number of approved FPs and the volume of finance allocated per GCF region (Table 1). Although the Asia–Pacific region claims the most proposals (120, or 42 per cent), followed closely by Africa (118, or 41 per cent), their per-project finance is not necessarily the highest. For instance, the Latin America and the Caribbean has fewer proposals (72), yet its total allocation of USD 3.8 billion translates to the highest level of per-project funding at approximately USD 53 million per proposal. Africa follows closely at around USD 51 million per proposal, while the Asia–Pacific region averages approximately USD 46 million per proposal. Eastern Europe, with only 18 proposals (6 per cent), has the lowest per-project funding at approximately USD 36 million. These differences suggest that factors beyond the number of proposals – such as project size, complexity, and region-specific funding priorities – can significantly influence how finance is ultimately allocated.

Table 1. Distribution of the GCF portfolio by region

GCF REGION	NUMBER OF FPS	FINANCE VOLUME (USD MIL)	SHARE OF TOTAL
Africa	118 (41%)	6,077	38%
Asia–Pacific	120 (42%)	5,554	34%
Eastern Europe	18 (6%)	654	4%
Latin America and the Caribbean	72 (25%)	3,813	24%
Total	**286***	**16,098**	**100%**

Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including all projects approved through B.40.
Notes: * The number of FPs does not aggregate into a total due to some approved proposals covering several regions. For projects across multiple regions, the regional distribution of funding is calculated based on the country-level financing allocations provided in the GCF Secretariat's public API. This methodology ensures that each region is credited with a more accurate funding allocated to countries within that region, rather than assigning the entire project budget to a single region.

14. Notable changes from previous data from the last evaluability study in 2023 are as follows:

(a) Latin America and the Caribbean's portfolio has grown from 62 to 72 proposals, with funding increasing from USD 3.2 billion to USD 3.8 billion.

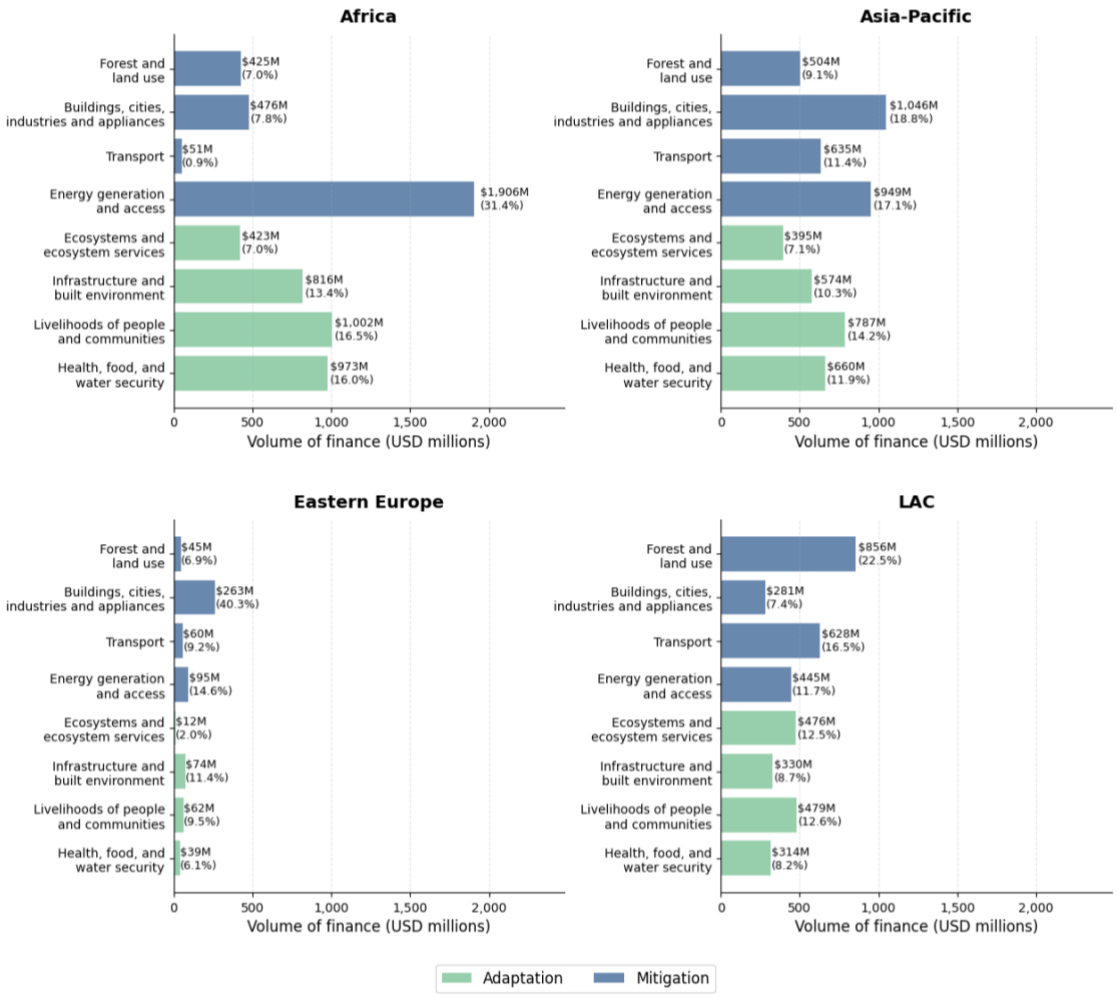
(b) The Asia–Pacific region now slightly exceeds Africa in the number of proposals (120 versus 118).

(c) Africa leads in total finance volume with USD 6.1 billion (38 per cent of the total), followed by Asia–Pacific with USD 5.6 billion (34 per cent), while Latin America and the Caribbean ranks third with USD 3.8 billion (24 per cent).

15. The breakdown of the GCF's investments across result areas offers valuable insights into its allocation strategies and priorities (see Figure 3). Among adaptation interventions, the largest investment – USD 2.3 billion – is directed towards enhancing the livelihoods of people and communities. This allocation reflects the GCF's commitment to strengthening community resilience as a key strategy for mitigating climate change impacts. Additionally, significant investments support health, food, and water security, with USD 2 billion allocated, representing a notable increase from the previous analysis, and USD 1.8 billion directed towards infrastructure and the built environment. These allocations highlight the Fund's focus on strengthening essential services and resilience mechanisms in vulnerable regions.

16. Regarding mitigation interventions, GCF investments are predominantly directed towards energy generation and access, with a substantial sum exceeding USD 3.3 billion. This significant investment reflects the Fund's commitment to advancing renewable energy sources and improving access to clean energy technologies, crucial steps in reducing greenhouse gas emissions.

Figure 3. Result areas finance by region



Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including all projects approved through B.40.
Note: Result area financing by region is derived from the GCF Secretariat's public API by applying a two-step allocation: (1) regional distribution based on country-level financing data within each project, and (2) result area distribution according to each project's specified investment weights. This ensures a more accurate representation of both geographical and thematic allocation of GCF resources. LAC stands for Latin America and the Caribbean.

2.4 Summary of approved projects in 2024

17. This section presents a summary of the 44 projects approved by the GCF in 2024, during Board meetings B.38, B.39, and B.40. These projects represent a total approved amount of USD 2,551 million, accounting for approximately 16 per cent of the overall portfolio of 286 projects. The set includes both FPs and SAP projects, while excluding those that were withdrawn, never submitted, or whose approvals lapsed.

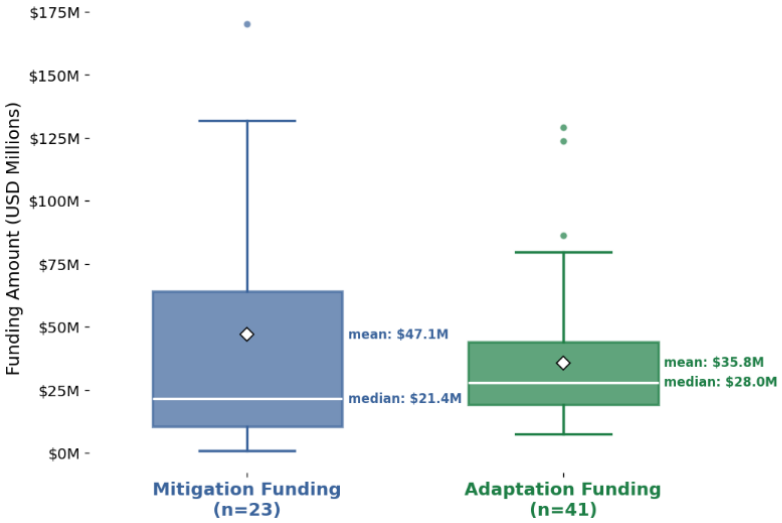
2.4.1 Portfolio by theme – mitigation versus adaptation funding

18. The distribution of funding by action type in the subset of 44 projects approved at Board meetings from B.38 to B.40 reveals clear differences between mitigation and adaptation projects, as shown in Figure 4. Between these three Board meetings, 21 adaptation projects were approved; 20 cross-cutting; and 3 purely mitigation projects were approved. Splitting cross-cutting projects into mitigation and adaptation, for mitigation, the median project size is USD 21.4 million, while the mean reaches USD 47.1 million, reflecting the influence of a few large-scale projects that raise the average above the median (outliers up to USD 170 million). For adaptation, the median is higher (USD 28.0 million) and the mean lower (USD 35.8 million), with a tighter distribution and a maximum of USD 129 million. This pattern suggests that adaptation financing is more evenly distributed, with a bias towards mid-scale projects, whereas mitigation displays a more heterogeneous profile, combining smaller interventions with a limited number of high-value operations.

19. When compared with the overall portfolio of 286 projects approved up to B.40, the subset of 44 projects displays a different profile from the aggregate (see Figure 2). In the full portfolio, mitigation accounts for 54 per cent of funding, with higher medians and means (USD 28.3 million and USD 53.9 million, respectively), while adaptation accounts for 46 per cent, with lower values (USD 25.0 million and USD 34.2 million). By contrast, the 44 projects show a higher share of adaptation (57.6 per cent) and a higher median size of adaptation projects (USD 28.0 million compared to 25.0 million in the global portfolio). Mitigation, in turn, declines to 42.4 per cent of total funding in this subset, with a lower median (USD 21.4 million), pointing to a concentration in smaller-scale projects, although a few large approvals keep the mean elevated.

20. It should be noted that cross-cutting projects were proportionally allocated to mitigation and adaptation themes according to their stated funding shares. These are: FP225, FP228, FP230, FP233, FP235, FP236, FP238, FP239, FP240, FP241, FP242, FP243, FP248, FP249, FP253, FP254, SAP037, SAP043, SAP045 and SAP047. Because of this apportionment, the number of components analysed (n=23 mitigation and n=41 adaptation) exceeds the total number of approved projects (44).

Figure 4. Climate finance allocation by action type for projects approved in 2024



Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including projects approved from B.38 to B.40.
Note: Cross-cutting projects are represented in both mitigation and adaptation categories according to their result area investment weights as specified in the GCF Secretariat's public API data. Therefore, the sum of projects across both categories exceeds the total number of projects approved from B.38 to B.40 (44), as cross-cutting projects with both components are counted proportionally in each category.

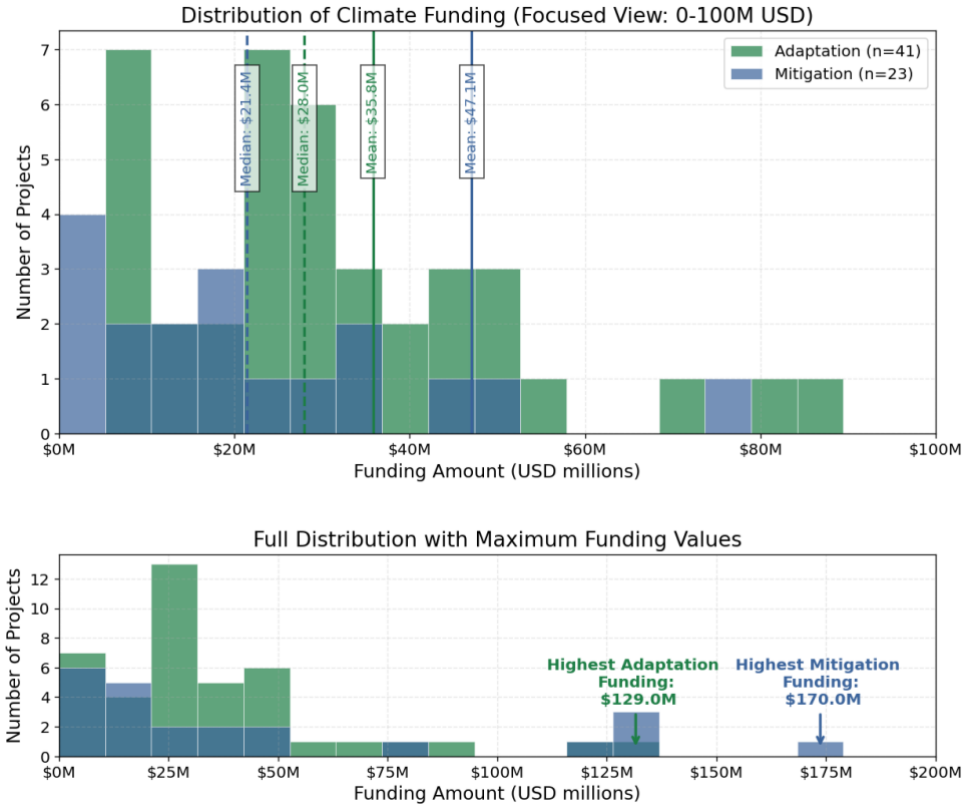
21. The analysis of funding distribution in the subset of 44 projects approved at Board meetings from B.38 to B.40 provides two complementary perspectives, presented in Figure 5.

22. The top panel (USD 0–100 million) shows that the vast majority of projects are concentrated within this range: 95.1 per cent of adaptation components and 78.3 per cent of mitigation components fall below this threshold. Adaptation is strongly clustered between USD 20 and 45 million, reflecting a more homogeneous and consistent pattern. Mitigation, by contrast, displays greater dispersion, combining smaller projects with a limited number of higher-value cases.

23. The bottom panel, which incorporates the full distribution, highlights the maximum values within the subset: USD 129 million for adaptation and USD 170 million for mitigation. Even considering these outliers, all 44 projects remain below the ~USD 200 million threshold, placing them in the lower- and mid-funding tiers of the Fund. This contrasts with the overall portfolio of 286 projects, where several projects exceed USD 300 million. The absence of such high projects in this subset does not necessarily indicate lower overall resourcing but rather reflects the nature and characteristics of the approved initiatives, with a predominance of mid-scale projects.

24. From a strategic perspective, this pattern suggests that Board decisions during this period favoured a more diversified and compact set of projects, particularly in adaptation, where the mid-range of funding clearly dominates. For mitigation, the presence of a few larger projects alongside a broad base of mid-sized interventions reinforces the notion of a more heterogeneous and fragmented portfolio, contrasting with the “high projects” that in earlier years had substantially increased the averages of the overall portfolio.

Figure 5. Climate finance allocation between adaptation and mitigation for projects approved in 2024



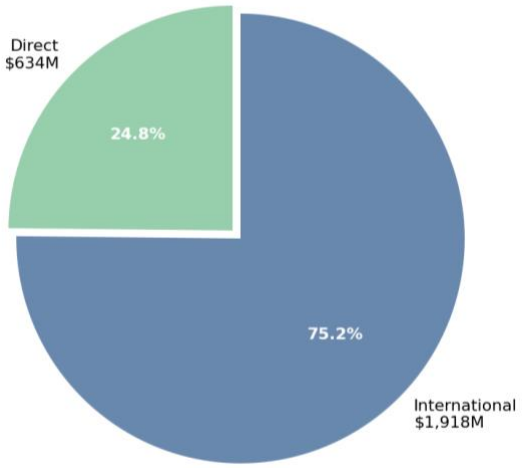
Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including projects approved from B.38 to B.40.
Note: Cross-cutting projects are disaggregated into mitigation and adaptation components according to their result area investment percentages provided in the GCF Secretariat's public API, allowing a more accurate representation of finance allocation by climate action type.

25. The subset of 44 projects departs from the Fund's historical pattern, with adaptation gaining prominence and exhibiting greater consistency within mid-scale tiers, while mitigation appears more dispersed and without exceptionally large-scale operations. As a result, the portfolio reflects a more balanced funding structure, reducing dependence on singular large-scale operations.

2.4.2 Funding by access modality (IAEs versus DAEs) in 2024

26. Within the 44 projects approved at Board meetings from B.38 to B.40, funding allocation is markedly skewed towards international accredited entities (IAEs): USD 1,916 million (75.2 per cent) across 30 projects, compared with USD 633 million (24.8 per cent) across 14 projects for direct access entities (DAEs). These figures suggest IAEs oversee both a greater volume and larger-scale operations (see Figure 6).

Figure 6. Climate finance allocation by access type for projects approved in 2024



Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including projects approved from B.38 to B.40.

27. IAEs remain the primary channels for large-scale financing within the Fund, while DAEs, though representing a smaller share in funding terms, continue to play a critical role in advancing the Fund's goals of direct access, national capacity-building and local ownership. This distribution highlights the coexistence of two complementary approaches: one emphasizing scale and financial reach through IAEs, and another fostering local ownership through DAEs.

2.4.3 Regional distribution and result areas in 2024

28. The 44 projects approved between the B.38 to the B.40 translate into 50 regional allocations because multi-region initiatives are counted in multiple geographic categories; specifically, FP225, FP253, and FP254 span two or more regions (see Figure 7). The cases that explain this difference are FP225 (Eastern Europe and Central Asia and Asia–Pacific), FP253 (Eastern Europe and Central Asia, Africa, and Asia–Pacific), and FP254 (Eastern Europe and Central Asia, Africa, Asia–Pacific, and Latin America and the Caribbean). In Asia–Pacific, with a total of USD 1,029 million, mitigation is concentrated in transport as the highest value (USD 194 million), while the lowest allocation corresponds to forest and land-use (USD 103 million). For adaptation, the highest allocations are found in livelihoods of people and communities (USD 156 million) and health, food, and water security (USD 155 million), compared to ecosystems and ecosystem services, which register the lowest value (USD 45 million). This pattern reflects a dual strategy: large-scale investments in transport and energy combined with programmes aimed at community resilience.

29. In Africa, which mobilized USD 769 million, adaptation clearly dominates. The highest allocation goes to health, food, and water security (USD 240 million), followed by livelihoods of people and communities (USD 223 million), while the lowest corresponds to infrastructure and built environment (USD 68 million). In mitigation, buildings, cities, industries and appliances account for the highest value (USD 72 million), while transport records only USD 1 million. This structure confirms Africa's role as the priority region for adaptation, in line with its high structural vulnerability.

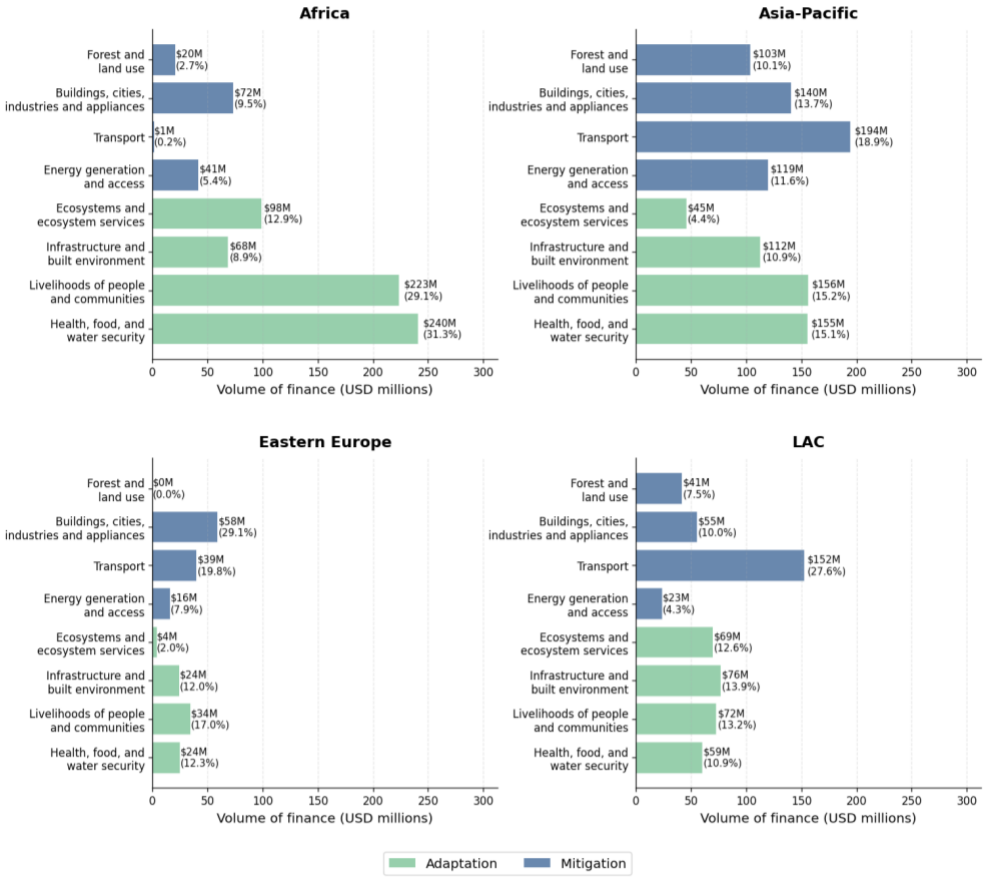
30. In Latin America and the Caribbean (USD 552 million), the profile combines mitigation and adaptation in a balanced way. In mitigation, transport leads with USD 152 million, while the lowest value is found in energy generation and access (USD 23 million). In adaptation, the highest allocation corresponds to infrastructure and built environment (USD 76 million), while

health, food, and water security records the lowest value (USD 59 million). This configuration reflects a dual priority: advancing the transition to low-emission transport systems and strengthening territorial and social resilience.

31. In Eastern Europe and Central Asia, with USD 202 million, mitigation shows its highest allocation in buildings, cities, industries and appliances (USD 58 million), while forest and land-use records no resource invested in this period of time (USD 0 million). In adaptation, livelihoods of people and communities account for the highest value (USD 34 million), compared to ecosystems and ecosystem services, which receive the lowest (USD 4 million). This pattern confirms a profile oriented towards energy efficiency and industrial infrastructure, with adaptation more contained than in other regions.

32. When compared with the overall portfolio of 286 projects, the results confirm consistency in thematic priorities: transport, energy, and building efficiency dominate mitigation, while health-food-water and livelihoods lead adaptation. However, unlike the full portfolio, where several projects exceed USD 300 million, this subset focuses on mid-scale operations, with a marked emphasis on resilience and food security. This suggests that recent Board approvals consolidate adaptation as a central axis, while mitigation remains anchored in emblematic sectors but with less prominence of mega-projects.

Figure 7. Climate finance by result area and region for projects approved in 2024



Source: GCF project API data as of December 31, 2024 (accessed July 21, 2025), including projects approved from B.38 to B.40.
Note: Result area financing by region is derived from the GCF Secretariat's public API by applying a two-step allocation: (1) regional distribution based on country-level financing data within each project, and (2) result area distribution according to each project's specified investment weights. This ensures a more accurate representation of both geographical and thematic allocation of GCF resources for the year 2024.

III. Methods overview

33. This chapter describes the methodological framework used to assess the evaluability of GCF FPs. The central question is: **To what extent are approved projects likely to credibly generate and report the results they claim?**

34. The analysis draws exclusively on desk reviews of FPs approved by the GCF Board. It assesses whether the information and proposed monitoring and evaluation systems described in these proposals are fit for purpose in fulfilling the GCF's objectives.

35. While the analysis recognizes the proposal designs evolve during implementation, the guiding principle is to evaluate what has been formally submitted for Board approval. The study also provides constructive insights on how proposals could be strengthened to better demonstrate credible and measurable results.

36. This section is structured as follows. Subsection 3.1 introduces the four analytical dimensions used to assess evaluability. Subsection 3.2 explains the stoplight framework that translates assessments into standardized risk ratings, while subsection 3.3 outlines the construction of the composite evaluability risk index. Subsection 3.4 describes how each dimension is operationalized through specific assessment criteria and guiding questions. Subsection 3.5 discusses the interpretation and limitations of the approach.

3.1 Analytical framework – four lenses to assess evaluability

37. The assessment is organized around four analytical dimensions or "lenses", that determine whether a proposal can credibly inform its intended results.

(a) **Theory of change** – whether proposals articulate a clear and coherent causal pathway, identify assumptions and unintended consequences, and draw on supporting evidence.

(b) **Potential for measuring causal change** – whether proposals outline credible evaluation strategies, including counterfactual logic, robust monitoring and evaluation (M&E) frameworks, and adequate resources for assessing causal impacts.

(c) **Implementation fidelity and investment criteria** – whether targeting and eligibility criteria are clearly defined, implementation risks are recognized, and the proposal demonstrates alignment with the GCF's six investment criteria: impact potential, paradigm shift potential, sustainability, country ownership, effectiveness and efficiency.

(d) **Data-collection and reporting credibility** – whether M&E plans are adequate, baseline data are available or planned, and indicators are clearly defined to measure progress and causal change.

38. These four dimensions are consistently applied throughout the analysis and are referred to as A–D in section IV to structure the stoplight results.

3.2 Building a stoplight – stoplight framework

39. Each proposal is evaluated against these four dimensions using a stoplight risk framework (see Box 1), which classifies the likelihood that the proposal meets the assessment criteria as low, medium, or high risk.

(a) Low risk: Strong likelihood of meeting assessment criteria; clear, complete and credible information.

(b) Medium risk: Partial or unclear information, requiring information; some aspects require clarification or strengthening.

(c) High risk: Weak or missing information; high likelihood of not meeting the assessment criteria.

40. This risk-based framework provides a standardized and comparable way of assessing evaluability "at entry". It focuses on projected success probabilities rather than observed outcomes, acknowledging that proposals may evolve as implementation progresses.

Box 1. Why use a risk framework?

The stoplight assessments associated with each GCF-funded proposal are constructed based on the information provided within the proposals. The proposals submitted to the GCF do not include every minute detail about the proposed project or programme. The GCF recognizes that the information in the proposals may be further adjusted based on feedback from the GCF, resulting from the evolving needs of target recipients or ongoing M&E efforts during implementation. Because the proposals are used as input for evaluating proposed projects or programmes, the project's quality vis-à-vis each stoplight criterion cannot be evaluated with absolute certainty before implementation. However, projects and programmes can be evaluated in terms of the *likelihood* they will meet each stoplight criterion based on the information in the proposal. Because the assessments gauge probabilities of success rather than the observed performance against the stoplight criteria, a risk framework provides a useful assessment tool. As described above, a project is rated as "high risk" for a given stoplight criterion when there is a high probability that the project described in the proposal will not adequately perform relative to that criterion. Alternatively, a "low risk" rating corresponds to a low probability of poor performance against a given criterion. This framework recognizes that our assessments are not based on observed progress but on the projected success of the proposed projects and programmes.

3.3 Constructing evaluability risk index

41. To summarize overall risk profiles, stoplight ratings are converted into numerical scores – 1 for low, 3 for medium and 5 for high – and aggregated across dimensions to create a composite evaluability risk index. Each subquestion receives equal weight in the composite calculation; given that all four dimensions contain an equal number of subquestions, the dimensions themselves are equally weighted. Unknown responses, which occur infrequently, are excluded from the aggregation.

42. The composite risk index is particularly useful for **comparing groups of proposals** rather than assessing individual projects. Lower index values indicate stronger proposal quality and greater evaluability readiness. This approach allows systematic comparisons across proposals, portfolio categories (for example, IAEs versus DAEs, or SAP versus PAP), and time periods (such as initial resource mobilization (IRM), GCF-1, and GCF-2).

3.4 Operationalization of criteria

43. Each of the four analytical dimensions is further broken down into specific evaluative questions and rating rules, designed to guide consistent application of the framework.

3.4.1 Theory of change and discussion of causal pathways

44. We use the following questions and rating rules to assess the quality of the theories of change and causal pathways discussed in the FPs.

1) What is the quality of the (implicit or explicit) theories of change and programme logic? (See the Annex I for a theory of change checklist)
   a) Low risk. Theory of change is well articulated.
   b) Medium risk. Logic framework or theory of change is present but needs some clarification. (Missing information is specified.)
   c) High risk. Logic framework or theory of change either does not exist, or it exists but relies on unverified assumptions or is missing critical details about implementation and/or causal pathways. (Missing information is specified.)
   d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating the theory of change.
2) Are unintended consequences referred to and identified robustly in the programme theory of change and/or in the surrounding literature reviews?
   a) Low risk. Unintended consequences are well articulated. (These are drawn from discussion of the theory of change.)
   b) Medium risk. Unintended consequences are discussed but need some clarification. (Missing information is specified.)
   c) High risk. Unintended consequences are not discussed and are potentially very large, given the programme design. (Missing information is specified.)
   d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating how it addresses unintended consequences.
3) Are causal pathways clearly identified and discussed? (This is discussed in the context of the theory of change and the credibility and feasibility of the pathways.)
   a) Low risk. Causal pathways are well articulated and supported with credible evidence.
   b) Medium risk. Causal pathways are described or implied, but the proposed links need some clarification about the assumptions on which they rely. (Missing information is specified.)
   c) High risk. The causal pathways implied in the proposal do not have a clear description and/or are based on unfounded assumptions.
   d) Unclear. Insufficient or ambiguous information in the proposal prevents evaluating the proposed causal pathways adequately.
4) How robust are the causal linkages (implicit or explicit) and are they well informed by high-quality evidence?
   a) Low risk. Causal linkages are well articulated and are well informed by high-quality evidence.
   b) Medium risk. Causal linkages are discussed but need clarification and/or support by additional high-quality evidence. (Missing information is specified.)
   c) High risk. Causal linkages are either not discussed or implied but lack any foundation in credible evidence. (Missing information is specified.)

d) Unclear. Insufficient or ambiguous information in the proposal prevents evaluating the proposed causal pathways.

5) Is good-quality evidence cited to discuss the efficacy of causal linkages?

a) Low risk. Evidence is of good quality and well articulated.

b) Medium risk. Evidence is used but needs some clarification. (Missing information is specified.)

c) High risk. Evidence is not discussed, or the quality of the evidence cited is inferior. (Missing information is specified.)

d) Unclear. The quality of the evidence cited to discuss the efficacy of causal linkages is unclear.

3.4.2 Potential for measuring causal change and evaluability

45. We ask the following questions to determine if causal change can be attributed to the GCF programme/GCF investment through impact evaluation (IE).

1) Does the proposal design allow for credible reporting of causal change?

a) Low risk. The proposal design allows for credible evaluation methods to be used to report casual change.

b) Medium risk. More details are needed to determine what could be a relevant comparison group or if there are feasible options to create comparison groups.

c) High risk. There does not appear to be a way to create a comparison group.

d) Unclear. There is not enough information to determine whether a credible measurement of causal change is possible.

2) To what extent are included requirements for M&E adequate and able to cover the costs of undertaking high-quality IEs?

a) Low risk. Requirements for M&E are likely adequate to cover the costs of a high-quality evaluation.

b) Medium risk. Requirements for M&E are specified but are likely insufficient to support a high-quality IE.

c) High risk. Requirements for M&E are not specified or cannot be determined from the information provided.

d) Unclear. Information about the M&E requirements is ambiguous, making assessing this information impossible.

3) Activities included in the proposal focus on “economic analyses” and “overall M&E” – are these sufficient for high-quality, credible evaluations?

a) Low risk. Both are specified and are of high quality.

b) Medium risk. Both are specified but are of low quality. (Missing information is specified.)

c) High risk. Only one is specified or neither is specified. (Missing information is specified.)

d) Unclear. Insufficient or ambiguous information in the proposal prevents adequate evaluation of the quality of proposed economic analyses and M&E activities.

4) Are methods for measuring attributable causal changes (outcomes, impact or other) discussed?

   a) Low risk. Measurement of attribution is well articulated.

   b) Medium risk. Measurement of attribution is discussed and/or the need for causal impact measurement is acknowledged, but strategies for doing so are not well articulated. (Missing information is specified.)

   c) High risk. Measurement of causal impact attribution is not discussed and/or the need for causal impact measurement is not acknowledged.

   d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating any proposed methods for measuring attributable causal changes.

5) Are there potential areas of bias that are likely to creep in?

   a) Low risk. There is a low risk of bias considering the proposed method of evaluating causal impact.

   b) Medium risk. There is a medium risk of bias considering the proposed method of evaluating causal impact. We specify what could lead to biases.

   c) High risk. There is a high risk of bias. The proposal either does not discuss a strategy for causal IE, or the strategy that is discussed has a high risk of producing unbiased impact estimates.

   d) Unclear. Cannot judge the likelihood of bias due to insufficient information.

3.4.3 Implementation fidelity and performance against investment criteria

46. We ask the following questions to determine if implementation and performance are likely to fit with the investment criteria.

1) Are eligibility and targeting criteria well articulated in submitted documents?

   a) Low risk. Eligibility and targeting criteria are well articulated.

   b) Medium risk. Eligibility and targeting criteria are discussed but need some clarification. We specify the missing information.

   c) High risk. Eligibility and targeting criteria are either not discussed or are discussed but do not appear feasible, given the programme design. (Missing information is specified.)

   d) Unclear. Insufficient or ambiguous information in the proposal prevents evaluating eligibility and targeting criteria adequately.

2) Is there adequate and reliable information included in the proposal regarding implementation fidelity?

   a) Low risk. Implementation fidelity appears to be strong.

   b) Medium risk. There is a medium level of risk related to implementation fidelity. Some risks to implementation fidelity need to be addressed. (Missing information is specified.)

   c) High risk. There is a high level of risk related to implementation fidelity. Substantial risks need to be addressed. We specify the missing information.

   d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating the information regarding implementation fidelity.

3) To what extent is impact potential identifiable and measurable in the proposal?

   a) Low risk. Impact potential is well articulated in the proposal and appears to be measurable using high-quality methods.

   b) Medium risk. Impact potential is specified but needs some clarification. We specify the missing information. Impact potential is measurable, but high-quality methods may not be feasible given the programme design.

   c) High risk. Impact potential is specified, but it relies on significant assumptions that are not verified, and/or impact indicators are vaguely described. Measurement and evaluation potential appears to be low.

   d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating the impact potential description and the feasibility of high-quality impact measurement.

4) To what extent is paradigm shift potential identifiable and measurable in the proposal?

   a) Low risk. Paradigm shift potential is well articulated in the proposal and appears to be measurable using high-quality methods.

   b) Medium risk. Paradigm shift potential is specified but needs some clarification. (Missing information is specified.) Paradigm shift potential is measurable, but high-quality methods may not be feasible given the programme design.

   c) High risk. Paradigm shift potential is specified, but it relies on significant assumptions that are not verified and/or paradigm shift indicators are vaguely described. Measurement and evaluation potential appears to be low.

   d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating the description of the potential paradigm shift and the feasibility of high-quality measurement.

5) How well are other GCF investment criteria informed, and are these measurable and verifiable with high credibility and quality?

   a) Low risk. Other investment criteria are likely to be credible.

   b) Medium risk. Other investment criteria have some limitations. (Missing information is specified.)

   c) High risk. Other investment criteria are not likely sufficient. We specify the missing information.

   d) Unclear. The credibility of other investment criteria cannot be determined from the information provided.

3.4.4 Data-collection and reporting credibility

47. We ask the following questions to determine if data-collection and reporting will likely be of good quality.

1) Are current reporting requirements sufficient for regular M&E?

   a) Low risk. Reporting for M&E is well articulated.

   b) Medium risk. Reporting for M&E is discussed but needs some clarification. We specify the missing information.

   c) High risk. Reporting for M&E is discussed, but it is insufficient for credible and useful M&E. We specify the missing information.

d) Unclear. The quality of reporting plans for M&E cannot be determined from the information provided.

2) How feasible is it to measure and report credibly on the progress of investment criteria, given M&E plans, budget and indicators for investment criteria?

a) Low risk. M&E and reporting plans have a high potential to measure progress on investment criteria.

b) Medium risk. M&E and reporting plans are discussed but are likely not of high enough quality or backed by sufficient resources to measure progress against investment criteria adequately.

c) High risk. M&E and reporting plans related to progress on investment criteria are not well articulated and/or clearly lack the resources needed to measure progress.

d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating the potential for the project to monitor and report on progress associated with investment criteria credibly.

3) To what extent did the proposal provide additional impact indicators beyond those proposed by the GCF? Can the proposal's indicators measure the magnitude of causal change?

a) Low risk. Indicators and measurements are well defined and can be used to measure impact.

b) Medium risk. Indicators and measurements lack specificity, and measuring impact using the specified indicators may be challenging.

c) High risk. Indicators and measurements are vague and/or unclear. More detailed indicators are needed to measure impacts credibly.

d) Unclear. Insufficient information in the proposal to deduce the quality of indicators and measurements.

4) Have baseline data been collected and/or is there a requirement for this?

a) Low risk. Project will use baseline data, and the methods for collecting are well articulated.

b) Medium risk. Baseline data are discussed but need some clarification. Missing information to be specified.

c) High risk. Plans for collecting baseline data are not discussed despite a need to collect baseline data to inform an IE.

d) Unclear. Insufficient or ambiguous information in the proposal prevents adequately evaluating plans for collecting baseline data.

5) What is the potential quality of the data and are they suitable for IEs?

a) Low risk. Data to be collected will be of high quality.

b) Medium risk. Data are likely to be of good quality.

c) High risk. Data are likely low quality, or data-collection plans are not specified/unclear.

d) Unclear. Insufficient or ambiguous information in the proposal prevents evaluating the potential quality of data adequately.

3.5 Interpretation and limitations

48. The analysis is based entirely on proposal documentation submitted to the GCF. Because these proposals do not represent finalized implementation plans, the assessments should be interpreted as indicative of potential evaluability rather than observed performance.

49. Furthermore, while the GCF encourages robust results measurement, many of the attributes examined here (e.g., explicit causal frameworks, counterfactual logic, M&E budgeting) were not formal GCF requirements at the time of proposal submission. Accordingly, the results are not judgments of implementing entities' capacity but rather assessments of the information environment that supports credible measurement and reporting of results.

50. The combined use of the four analytical dimensions, the stoplight framework, and the composite risk index serves three primary purposes:

(a) Identify recurring patterns and weaknesses in proposal design.

(b) Highlight opportunities to strengthen evaluability and measurement systems.

(c) Generate evidence to inform future improvements in proposal quality and reporting credibility.

IV. Stoplight analysis – Results

51. In this section, we present the key findings from the stoplight analysis of the GCF project portfolio as of 31 December 2024. In 2024, at B.38, B.39 and B.40, 44 new projects were approved, amounting to USD 2,551 million in funding and representing 16 per cent of the overall portfolio of 286 projects approved up to B.40. The analysis is organized into four parts. First, we review trajectories of risk ratings over time. Second, we compare results before and after the introduction of the revised FP template. Third, we examine differences between IAEs and DAEs. Finally, we assess the proposals approved in 2024, comparing them with those from earlier replenishment periods.

52. As of B.40, the GCF has approved a total of 286 funded projects (both FP and SAP) after accounting for withdrawn and lapsed projects. The following assessment examines these 286 funded projects, comprising 121 projects approved during the IRM period,121 approved during the GCF-1 period, and 44 during GCF-2.[4]

4.1 Overall changes across the four assessment areas

53. The stoplight data are used to assess the quality and evaluability of the proposals at entry across four main areas as described earlier in the methods section. These areas are:

(a) Theory of change

(b) Potential for measurement of causal change and evaluability

(c) Implementation fidelity and performance against investment criteria

(d) Data-collection and reporting credibility

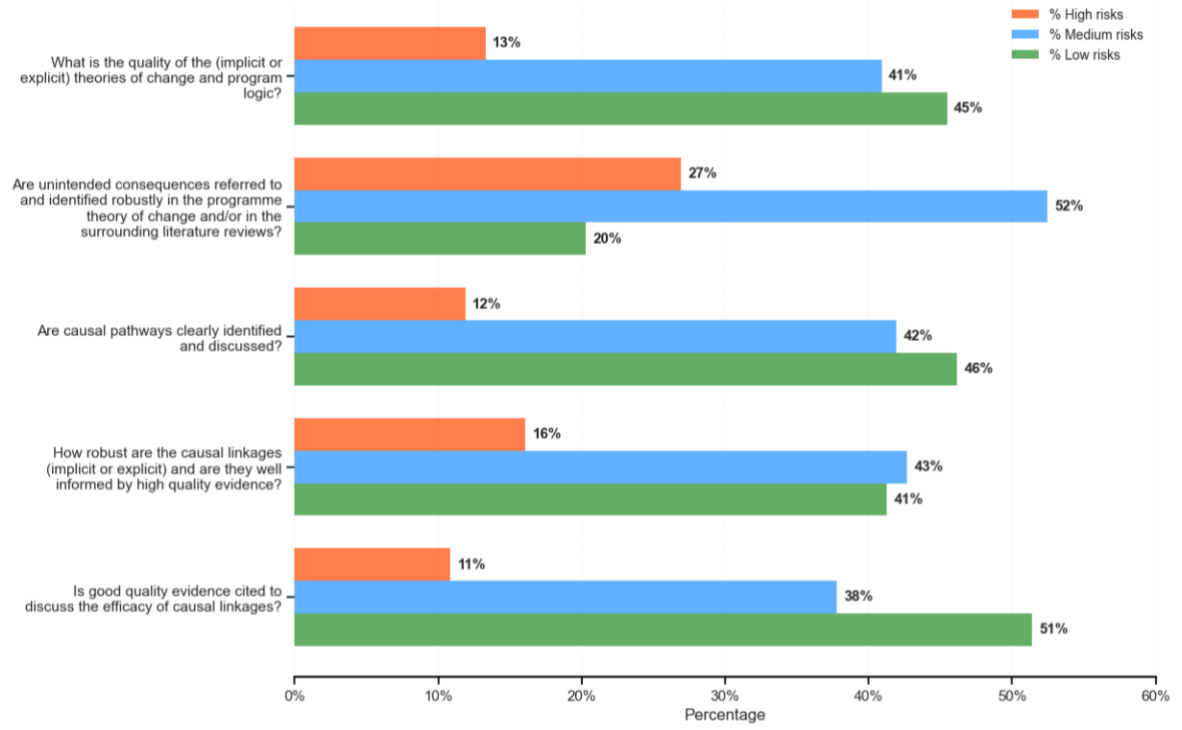
54. In category A, we assess the extent to which the pathways to impact are outlined using a theory of change or logic model. The assessment seeks to determine if the theory of change is explicit or implicit, to assess if the proposal identifies and/or is cognizant of the potential externalities of its requested financing, and to ascertain the robustness of the evidence cited to support the programme design (see Figure 8).

55. The best-performing criterion (most low-risk evaluations) relates to the use of good-quality evidence to discuss the efficacy of causal linkages. The identification of causal pathways, robustness of implicit and explicit causal linkages, and the quality of the theory of change show a more even distribution across risk levels, indicating variability in how well these elements are addressed in the GCF portfolio. The weakest area is the identification and referral of unintended consequences, with only 20 per cent of proposals at low risk and 27 per cent at high risk.

56. Overall, while proposals demonstrate clear strengths in evidence use and show moderate performance in theory of change quality and causal pathways, a consistent share of proposals still falls into the high-risk category (11–27 per cent across all criteria), indicating persistent heterogeneity in the clarity and robustness of project logic at the proposal stage.

[4] Data were last verified on 18 February 2025.

Figure 8. (A) Stoplight assessment of the theory of change

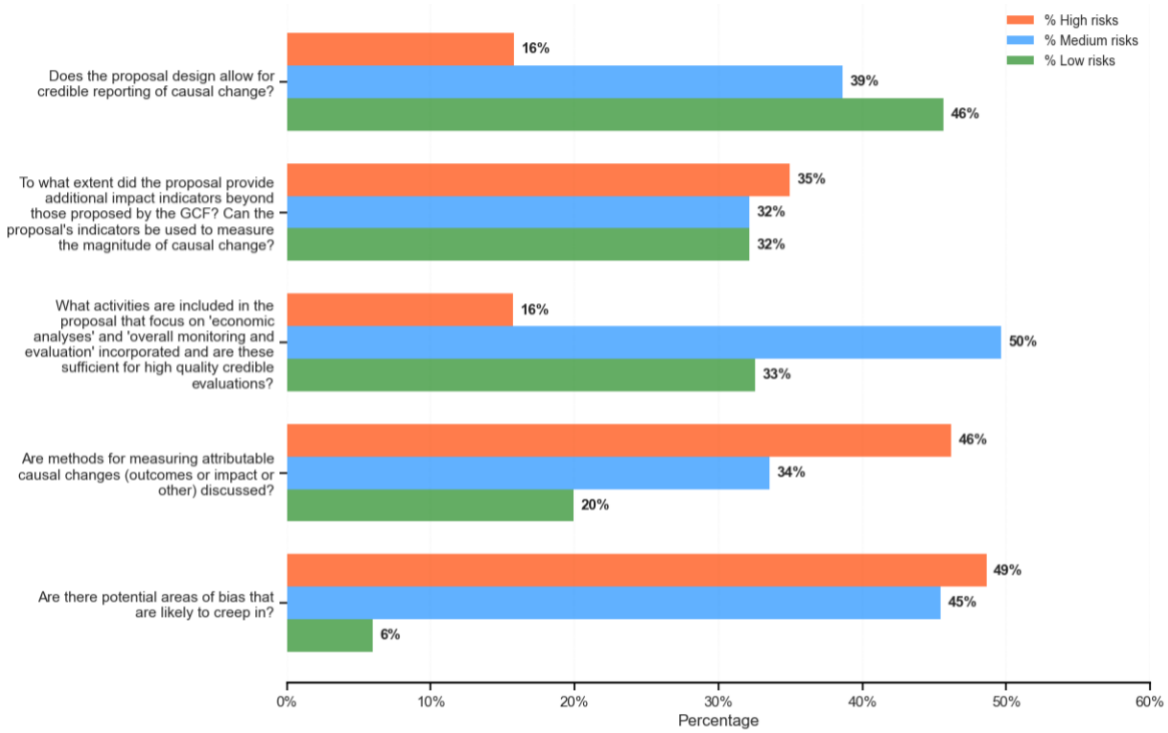


Source: FPs as of December 31, 2024.
Note: (n=286 projects); assessment and analysis performed by the authors.

57. In category B, we assess the ability to accurately measure, report and evaluate the economic impact and other changes due to the proposed activities. In other words, the focus is on whether the claimed causal effects can be credibly assessed, based on the FP's M&E plans (see Figure 9).

58. The proposals demonstrate a mixed distribution of risks across the assessed criteria. Lack of consideration of potential biases emerges as a particularly acute concern, with 49 per cent of the proposals assessed as high risk and a further 45 per cent as medium risk. Similarly, methods for measuring attributable causal changes also present substantial challenges, with 46 per cent of proposals rated as high risk, indicating continued challenges in how outcomes or impacts are evaluated. By contrast, the inclusion of economic analyses and M&E activities is less skewed towards high risk, with only 16 per cent scored as such, though half fall into the medium-risk category, indicating limited robustness. Design for credible reporting of causal change suggests relatively stronger performance compared to other criteria with 46 per cent classified as low risk. Meanwhile, the provision of additional impact indicators beyond those proposed by the GCF is evenly spread across risk categories, suggesting variability in the inclusion of additional impact indicators. Overall, these findings highlight persistent weaknesses in addressing bias and in methods for measuring causal change.

Figure 9. (B) Stoplight assessment of the potential for measurement of causal change and evaluability

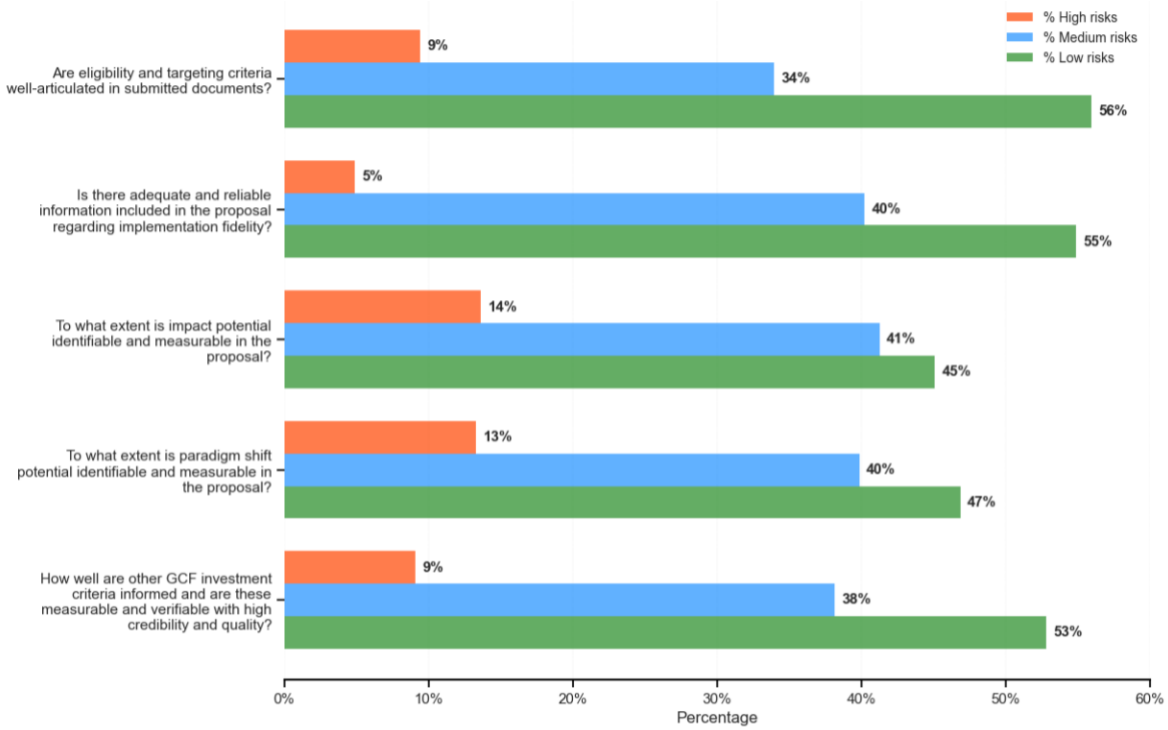


Source: FPs as of December 31, 2024.
Note: (n=286 projects); assessment and analysis performed by the authors.

59. In category C, we assess whether project activities are well targeted, examine the programme's performance against the GCF's investment criteria, and review the feasibility of the overall implementation plans. We also consider whether the proposal identifies relevant barriers to implementation and outlines appropriate measures for addressing them if they arise (see Figure 10).

60. The proportion of high risk is consistently lower, below 15 per cent across all the evaluation criteria, reflecting an overall good level of quality at the proposal stage. Strengths are most evident in articulation of eligibility and targeting criteria, providing information on implementation fidelity, and responsiveness to other GCF investment criteria, where over half of the proposals are rated low risk. By contrast, evaluating impact potential and paradigm shift potential shows greater heterogeneity, with a sizeable share of proposals in the medium-risk range, indicating that while these aspects are generally addressed, they remain areas where proposals could be further strengthened.

Figure 10. (C) Stoplight assessment of implementation fidelity and performance against investment criteria

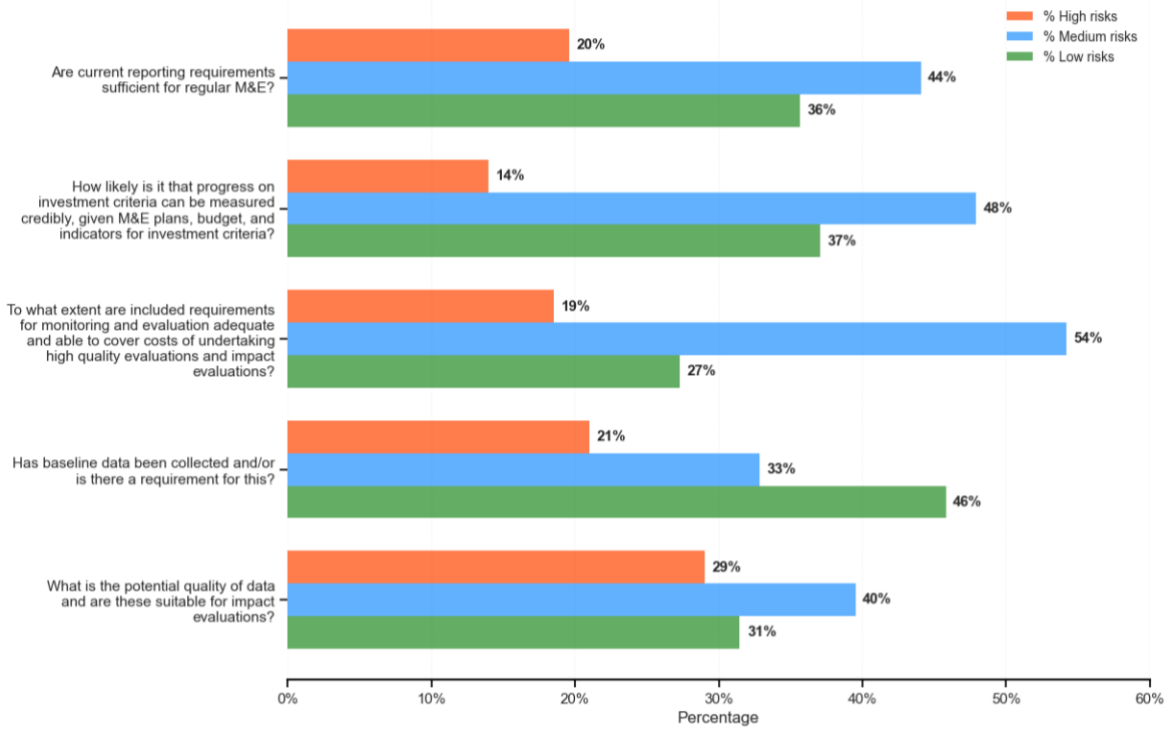


Source: FPs as of December 31, 2024.
Note: (n=286 projects); assessment and analysis performed by the authors.

61. Lastly, in category D, we assess whether the data-collection and reporting processes outlined in the proposals are rigorous enough to help identify the causal effects of the GCF investment (see Figure 11).

62. The criterion on data quality to assess impact shows moderate concerns, with 29 per cent proposals rated high risk, indicating potential issues with data quality and suitability. Availability of baseline data is more favourably assessed with approximately 46 per cent of proposals rated low risk, suggesting relatively better practices or clearer requirements in this area. At the same time, both baseline data availability and regular M&E reporting requirements show room for improvement, as approximately 20 per cent of proposals are still rated high risk. In terms of the robustness of the M&E requirement for high quality and IEs, the majority falls into medium risk, indicating that while there are efforts in budget planning to undertake evaluations, such cost consideration persists as a challenge. Credibility in reporting on investment criteria shows a substantial share of proposals at medium risk (48 per cent), reflecting ongoing challenges. At the same time, 37 per cent are rated low risk, suggesting a reasonable foundation in M&E planning and budget alignment.

Figure 11. (D) Stoplight assessment of data-collection and reporting credibility



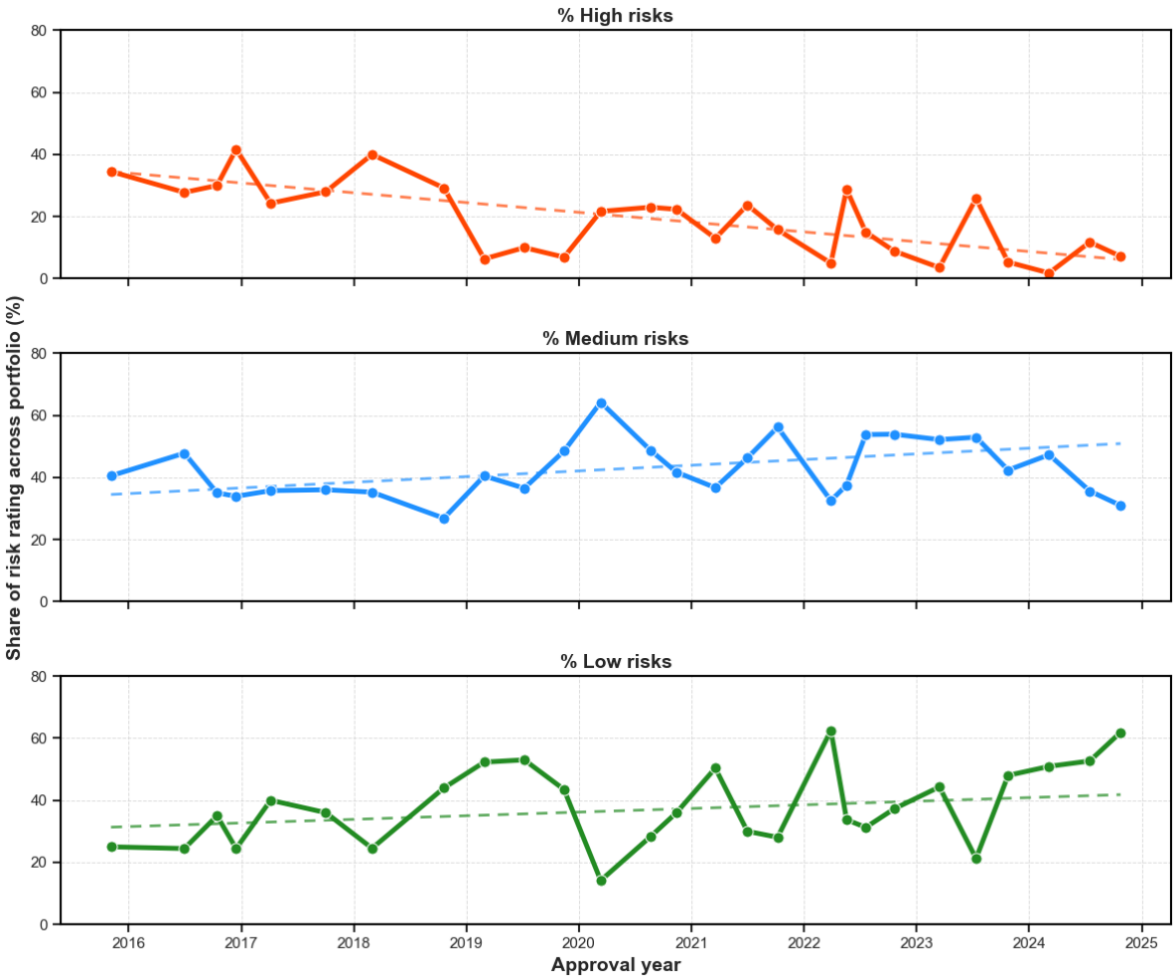
Source: FPs as of December 31, 2024.
Note: (n=286 projects); assessment and analysis performed by the authors.

4.2 Overall portfolio trajectories over time

63. Figure 12 tracks the evolution of proposal quality by illustrating the share of criteria rated as high, medium, or low risk across successive Board approvals. Rather than looking at single projects in isolation, this portfolio-level view highlights how the overall distribution of risks has shifted as the Fund has matured. Each data point reflects the proportion of risk ratings associated with proposals approved in a given Board meeting, while the dotted lines depict the underlying trends over the period.

64. At the initial data point (November 2015), more than 40 per cent of the criteria in project proposals were deemed high risk, whereas only about 20 per cent were rated as low risk. Over successive Board approvals, high-risk ratings have gradually declined, and medium- and low-risk ratings have gained prominence. This sustained trend, consistent with previous assessments, indicates that proposal design and evaluability have become progressively stronger as the Fund has matured.

Figure 12.. Trends in portfolio risk level over time (2016–2024)



Source: FPs as of December 31, 2024.
Note: Assessment and analysis by authors. Dotted lines show linear regression trends. These trend lines provide a simplified visualization of the overall direction of changeover the time period. A rising line indicates an increasing trend in risk exposure, and a declining line indicates a decreasing trend.

4.3 Assessment of portfolio: before and after the FP template change

65. In 2022, the GCF revised its FP template for both the standard PAP and SAP to align with the integrated results management framework, adopted at the B.29/01. This revision introduced clearer structures, standardized indicators, and streamlined requirements to enhance clarity, consistency, and alignment with results reporting. From B.32 (2022) onwards, all pipeline proposals, including resubmissions, have been required to use this updated format. The revised template strengthened FPs by restructuring the logical framework, standardizing core indicators, and ensuring greater comparability across projects.

66. The introduction of the revised FP template provides an opportunity to examine whether these changes, clearer guidance, more standardized requirements, and stronger results alignment, have translated into measurable improvements in evaluability. By comparing stoplight results from proposals assessed before and after the template change, this study identifies shifts in risk levels across key criteria and assesses whether recurring weaknesses have been effectively addressed.

67. Results show that the introduction of the revised FP template is associated with measurable improvements in proposal evaluability. The overall risk index decreased from 2.78 to 2.30 (based on a scale of 1 to 5), reflecting a moderate but meaningful shift towards lower risk. As shown in Table 2, improvements are evident across all four assessment criteria, with the largest gains in measurement of causal change and data-collection and reporting. The results suggest that the revised template has strengthened the clarity and consistency of proposal design, with the most notable improvements in the measurement of causal change and in data-collection and reporting systems.

Table 2. Comparison of the risk index between projects at B.31 and before, and at B.32 and after

Assessment criteria	B.31 and before (N=191)	B.32 and after (N=95)	Differences after template change
A. Theory of change	2.62	2.25	0.37↓
B. Measurement of causal change	3.34	2.60	0.74↓
C. Implementation fidelity	2.24	2.05	0.19↓
D. Data-collection and reporting	2.90	2.28	0.62↓
Overall	**2.78**	**2.30**	**0.48↓**

Source: FPs as of December 31, 2024.
Note: Assessment and analysis by authors.

If we focus on the share of proposals rated high risk, the comparison before and after the revised FP template shows clear improvements. Table 3 highlights some of the most substantial areas of reduction, led by impact indicators (-39.6 percentage points (pp)), bias identification (-33.4 pp), and data quality (-26.1 pp). Moderate gains are also evident in causal linkages. Overall, the FP template change has strengthened proposal quality in areas most critical to evaluability.

Table 3. Share of proposals rated high risk under selected assessment criteria: B.31 and before versus B.32 and after

Observation	Assessment criteria	B.31 and before (N=191)	B.32 and after (N=95)	Change (pp)*
		% high risks	% high risks	
Improved areas	D. To what extent did the proposal provide additional impact indicators beyond those proposed by the GCF? Can the proposal's indicators be used to measure the magnitude of causal change?	48.2%	8.6%	39.6 pp ↓
	D. Are there potential areas of bias that are likely to creep in?	59.7%	26.3%	33.4 pp ↓
	D. What is the potential quality of	37.7%	11.6%	26.1 pp ↓

Observation	Assessment criteria	B.31 and before (n=191)	B.32 and after (n=95)	Change (pp)*
		% high risks	% high risks	
	data and are these suitable for IEs?			
	B. Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	53.4%	31.9%	21.5 pp ↓
	A. How robust are the causal linkages (implicit or explicit) and are they well informed by high-quality evidence?	22.0%	4.2%	17.8 pp ↓
Worsened area	C. Is there adequate and reliable information included in the proposal regarding implementation fidelity?	4.7%	5.3%	0.6 pp ↑

Source: FPs as of December 31, 2024. Assessment and analysis by authors.
Notes: *Negative values indicate improvement (reduction in high risk), while positive values indicate worsening (increase in high risk). The pp change shows the magnitude of difference between the "B.31 and before" and "B.32 and after" assessment periods.

4.4 Assessment of portfolio: IAE versus DAE

68. This section compares the stoplight results of IAEs and DAEs to examine whether there are systematic differences in risk ratings across key assessment criteria. The overall risk index was 2.59 for IAEs and 2.69 for DAEs on a 1–5 scale, reflecting only a small difference of 0.10 points (Table 4). Slightly higher risks for DAEs are most evident in implementation fidelity (+0.17 points), with smaller gaps in theory of change and data-collection and reporting (+0.12 points each). Overall, the findings suggest that proposal quality at entry is broadly comparable between IAEs and DAEs, with only modest variation across assessment criteria.

Table 4. Comparison of the risk index between IAEs and DAEs

Assessment criteria	IAE (n=216)	DAE (n=70)	Differences
A. Theory of change	2.47	2.59	0.12
B. Measurement of causal change	3.08	3.16	0.08
C. Implementation fidelity	2.16	2.23	0.17
D. Data-collection and reporting	2.67	2.79	0.12
Overall	**2.59**	**2.69**	**0.10**

Source: FPs as of December 31, 2024.
Note: Assessment and analysis by authors.

69. To complement the overall comparison of risk indices, we now focus specifically on areas rated high risk (Table 5). The comparative assessment between proposals linked to DAEs and IAEs reveals notable differences, with DAEs showing higher risks in articulating causal pathways (10.7 points), establishing robust causal linkages (7.1 points), and meeting M&E reporting requirements (6.1 points). These findings suggest that although overall risk levels are broadly similar, DAEs face greater challenges in certain design-related dimensions that are critical for evaluability.

Table 5. Comparison of high-risk assessment criteria: IAE-led projects versus DAE-led projects

Observation	Assessment criteria	IAE (n=216)	DAE (n=70)	Differences
		% high risks	% high risks	
DAEs have higher risk than IAEs	A. Are causal pathways clearly identified and discussed?	9.3%	20%	10.7 pp
	A. How robust are the causal linkages (implicit or explicit) and are they well informed by high-quality evidence?	14.4%	21.4%	7.1 pp
	D. Are current reporting requirements sufficient for regular M&E?	18.2%	24.3%	6.1 pp
	C. To what extent is paradigm shift potential identifiable and measurable in the proposal?	12%	17.1%	5.1 pp
	A. What is the quality of the (implicit or explicit) theories of change and programme logic?	12.1%	17.1%	5 pp

Source: FPs as of December 31, 2024.
Note: *Negative values indicate improvement (reduction in high risk), whereas positive values indicate worsening (increase in high risk). The pp change shows the magnitude of difference between the GCF-1 and IRM assessment periods.

4.5 Assessment of portfolio: approved projects in 2024

70. This section reviews FPs approved in 2024 and compares their stoplight risk indices with those from the GCF-1 period (2020–2023). Proposals approved in 2024 show further improvements compared to GCF-1, with risk indices declining across all four assessment criteria (Table 6). The most notable gains are in measurement of causal change (-0.73), implementation fidelity (-0.67), and theory of change (-0.61), indicating stronger proposal quality in the early stages of GCF-2.

Table 6. Comparison of the stoplight assessment (risk index) between the IRM period (2015–2019), GCF-1 (2020–2023) and GCF-2 (2024–onwards)

Assessment criteria	IRM (N=121)	GCF-1 (N=121)	GCF-2 (2024) (N=44)	Change between GCF-1 and GCF-2
A. Theory of change	2.62	2.57	1.96	0.61↓
B. Measurement of causal change	3.40	3.07	2.34	0.73↓
C. Implementation fidelity	2.17	2.36	1.69	0.67↓
D. Data-collection and reporting	3.13	2.45	2.17	0.28↓
Overall	**2.83**	**2.61**	**2.04**	**0.57↓**

Source: FPs as of December 31, 2024.
Note: Assessment and analysis by authors.

71. Analysis of the changes between periods reveals:

(a) Theory of change: A decrease (-0.61) in risk score, suggesting clear improvement in project logical frameworks.

(b) Measurement of causal change: A significant decrease (-0.73) in risk score, indicating substantial improvement in how projects define, establish and measure cause–effect relationships.

(c) Implementation fidelity: A moderate decrease (-0.67) in risk score, pointing to better alignment between planned and actual implementation of project activities.

(d) Data-collection and reporting: A considerable decrease (-0.28) in risk score, demonstrating marked improvement in how projects gather, analyse and report data.

72. When interpreting these changes, shifts greater than 0.5 (as seen in theory of change) represent moderate improvements, whereas changes approaching or exceeding 1.0 (as seen in measurement of causal change, and data-collection and reporting) would indicate major shifts in project risk profiles. The overall trend suggests meaningful progress in risk reduction across all assessment criteria during the GCF-2 period.

V. Conclusion and discussion

73. The analysis confirms a consistent downward trend in high-risk ratings across GCF proposals since 2016, indicating gradual and sustained improvements in project design and evaluability as the portfolio has matured. This trend reflects the broader institutional learning process within the GCF, where iterative feedback, accreditation experience, and increased engagement with evaluation principles have contributed to higher-quality proposals.

74. The revised FP template introduced in 2022 is associated with lower risk indices across all assessment criteria, particularly in the areas of **measuring causal change** and **strengthening data-collection and reporting systems**. These improvements suggest that the template has enhanced the clarity, consistency and standardization of proposal submissions. However, the observed effect is difficult to attribute solely to this reform, as the cohort analysis between the IRM period and GCF-1 shows that risk reduction has been gradual and cumulative. The evidence therefore points to a broader process of portfolio strengthening over time, rather than a discrete policy shift.

75. Comparative analyses between **IAEs** and **DAEs** show broadly similar overall risk profiles, indicating convergence in proposal quality across access modalities. Nonetheless, DAEs tend to face greater challenges in several design dimensions, notably the articulation of **causal pathways**, the **robustness of causal linkages**, and the **adequacy of M&E requirements**. These differences may reflect variations in institutional experience, resource availability, and exposure to **Evaluation Standards**. Continued capacity support, including through readiness and preparatory programmes, could help address these gaps and further level the playing field between IAEs and DAEs.

76. The analysis of proposals approved in **2024 (early GCF-2 period)** provides additional evidence of continued improvement in evaluability compared with the GCF-1 period (2020–2023). The overall decline in risk indices across all four analytical dimensions, particularly in the measurement of causal change and data credibility, suggests that the Fund is entering a phase of greater methodological maturity. These gains are especially significant considering that GCF-2 marks the first programming cycle implemented entirely under the revised template and the integrated results management framework.

77. Despite these encouraging trends, the analysis also reveals persistent areas for improvement. Some proposals continue to **exhibit limited articulation of causal logic**, weak specification of counterfactual reasoning, and **insufficient budgeting for evaluation and data systems**. Addressing these gaps will be critical for enhancing **reporting credibility**, **learning value**, and **accountability for results** across the GCF portfolio. Strengthening the FPs evaluability, especially through early design-stage support, improved M&E guidance, and stronger integration of evaluative thinking within AEs, will enable the Fund to more effectively demonstrate the impact of its investments.

78. **Building on these findings, the IEU is advancing several complementary initiatives to institutionalize evaluability and strengthen evidence quality across the GCF portfolio.** One of these is the ongoing development of **Impact Evaluation Standards**, which will establish common principles and quality benchmarks for designing, conducting, and using impact evaluations within the GCF context. These standards aim to promote methodological rigour, transparency, and consistency in how causal impacts are measured and reported, ensuring alignment between project-level evaluations and the Fund's results architecture.

79. Another major initiative is the forthcoming **evaluation quality assessment (EQA)** process which will be introduced in 2026. The EQA provides an **independent assessment of the quality of interim and terminal evaluations conducted by AEs**. It systematically reviews the credibility, methodological soundness, and utility for learning of AE-led evaluations,

ensuring they comply with the GCF Evaluation Policy and adhere to the Evaluation Standards, as well as good practices set by the United Nations Evaluation Group. Results are synthesized semi-annually or annually to inform organizational learning, establish quality benchmarks, and provide targeted feedback to AEs. Unlike the evaluability assessment, which examines the potential for credible measurement of results at the design stage, the EQA reviews completed evaluations to assess their actual quality and credibility. Together, these initiatives reinforce the GCF commitment to a robust, evidence-based evaluation system that supports accountability, adaptive management, and continuous learning across its growing portfolio.

80. Looking ahead, **Impact Evaluation Standards** and the **EQA** process will play a pivotal role in further strengthening evaluative thinking throughout the project cycle, from design to implementation and post-completion review. These efforts will help ensure that the GCF's expanding portfolio is not only impactful in delivering climate results but also **credible, measurable, and grounded in trusted evaluation evidence**.

Annex I. Summary tables

Table A - 1. Stoplight assessment of the theory of change (by replenishment period)

Replenishment period	IRM (N=121)			GCF-1 (N=121)			GCF-2 (2024) (N=44)		
A. Theory of change	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
What is the quality of the (implicit or explicit) theories of change and programme logic?	38%	42%	20%	38%	51%	11%	89%	9%	2%
Are unintended consequences referred to and identified robustly in the programme theory of change and/or in the surrounding literature reviews?	38%	42%	20%	11%	53%	36%	0%	80%	20%
Are causal pathways clearly identified and discussed? (This is discussed in the context of the theory of change and the credibility and feasibility of the pathways.)	36%	44%	20%	47%	45%	8%	70%	30%	0%
How robust are the causal linkages (implicit or explicit) and are they well informed by high-quality evidence?	33%	42%	25%	40%	47%	13%	68%	32%	0%
Is good-quality evidence cited to discuss the efficacy of causal linkages?	50%	37%	13%	51%	37%	12%	57%	41%	2%

Table A - 2. Stoplight assessment of the theory of change (by type of access)

Access type	IAE (N=216)			DAE (N=70)		
A. Theory of change	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
What is the quality of the (implicit or explicit) theories of change and programme logic?	45%	43%	12%	47%	36%	17%
Are unintended consequences referred to and identified robustly in the programme theory of change and/or in the surrounding literature reviews?	22%	50%	27%	14%	60%	26%
Are causal pathways clearly identified and discussed? (This is discussed in the context of the theory of change and the credibility and feasibility of the pathways.)	46%	44%	9%	46%	34%	20%
How robust are the causal linkages (implicit or explicit) and are they well informed by high-quality evidence?	42%	44%	14%	39%	40%	21%
Is good-quality evidence cited to discuss the efficacy of causal linkages?	52%	37%	12%	50%	41%	9%

Table A - 3. Stoplight assessment of the theory of change (by changes of template: B.31 and before versus B.32 and after)

CHANGES IN PROPOSAL TEMPLATE	B.31 AND BEFORE (N=191)			B.32 AND AFTER (N=95)		
A. Theory of change	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
What is the quality of the (implicit or explicit) theories of change and programme logic?	36%	46%	17%	64%	31%	5%
Are unintended consequences referred to and identified robustly in the programme theory of change and/or in the surrounding literature reviews?	26%	45%	28%	8%	67%	24%
Are causal pathways clearly identified and discussed? (This is discussed in the context of the theory of change and the credibility and feasibility of the pathways.)	42%	42%	16%	55%	41%	4%
How robust are the causal linkages (implicit or explicit) and are they well informed by high-quality evidence?	36%	42%	22%	53%	43%	4%
Is good-quality evidence cited to discuss the efficacy of causal linkages?	51%	36%	13%	52%	42%	6%

Table A - 4. Stoplight assessment of the potential for measurement of causal change and evaluability (by replenishment period)

Replenishment period	IRM (N=121)			GCF-1 (N=121)			GCF-2 (2024) (N=44)		
B. Potential for measurement of causal change and evaluability	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Does the proposal design allow for credible reporting of causal change?	53%	34%	13%	31%	50%	20%	68%	20%	11%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high-quality IEs?	24%	50%	26%	25%	60%	16%	43%	50%	7%
Are activities included in the proposal that focus on "economic analyses" and "overall monitoring and evaluation", and are these sufficient for high-quality, credible evaluations?	31%	47%	21%	29%	55%	16%	50%	48%	2%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	15%	28%	57%	13%	43%	43%	52%	23%	25%
Are there potential areas of bias that are likely to creep in?	2%	28%	70%	7%	55%	37%	11%	66%	23%

Table A - 5. Stoplight assessment of the potential for measurement of causal change and evaluability (by type of access)

Access type	IAE (N=216)			DAE (N=70)		
B. Potential for measurement of causal change and evaluability	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Does the proposal design allow for credible reporting of causal change?	45%	38%	17%	47%	40%	13%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high-quality IEs?	28%	53%	19%	24%	59%	17%
Are activities included in the proposal that focus on "economic analyses" and "overall monitoring and evaluation", and are these sufficient for high-quality, credible evaluations?	34%	50%	16%	29%	54%	16%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	20%	34%	45%	19%	31%	50%
Are there potential areas of bias that are likely to creep in?	7%	44%	49%	1%	51%	47%

Table A - 6. Stoplight assessment of the potential for measurement of causal change and evaluability (by changes of template: B.31 and before versus B.32 and after)

Replenishment period	B.31 and before (n=191)			B.32 and after (n=95)		
B. Potential for measurement of causal change and evaluability	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Does the proposal design allow for credible reporting of causal change?	42%	41%	18%	54%	35%	12%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high-quality IEs?	25%	53%	21%	32%	56%	13%
Are activities included in the proposal that focus on "economic analyses" and "overall monitoring and evaluation", and are these sufficient for high-quality, credible evaluations?	29%	50%	21%	41%	52%	7%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	15%	32%	53%	31%	37%	32%
Are there potential areas of bias that are likely to creep in?	5%	35%	60%	7%	66%	26%

Table A - 7. Stoplight assessment of implementation fidelity and performance against investment criteria (by replenishment period)

REPLENISHMENT PERIOD	IRM (N=121)			GCF-1 (N=121)			GCF-2 (2024) (N=44)		
C. Implementation fidelity and performance against investment criteria	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Are eligibility and targeting criteria well articulated in submitted documents?	59%	26%	15%	48%	44%	8%	70%	30%	0%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	79%	18%	3%	30%	64%	7%	59%	36%	5%
To what extent is impact potential identifiable and measurable in the proposal?	50%	36%	14%	31%	52%	17%	70%	25%	5%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	47%	31%	22%	41%	50%	9%	61%	39%	0%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	44%	40%	16%	54%	40%	6%	75%	25%	0%

Table A - 8. Stoplight assessment of implementation fidelity and performance against investment criteria (type of access)

ACCESS TYPE	IAE (N=216)			DAE (N=70)		
C. Implementation fidelity and performance against investment criteria	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Are eligibility and targeting criteria well articulated in submitted documents?	57%	34%	10%	56%	36%	9%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	57%	39%	4%	49%	44%	7%
To what extent is impact potential identifiable and measurable in the proposal?	47%	38%	14%	39%	50%	11%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	49%	39%	12%	41%	41%	17%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	51%	39%	10%	59%	34%	7%

Table A - 9. Stoplight assessment of implementation fidelity and performance against investment criteria (by changes of template: B.31 and before versus B.32 and after)

Replenishment period	B.31 and before (n=191)			B.32 and after (n=95)		
C. Implementation fidelity and performance against investment criteria	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Are eligibility and targeting criteria well articulated in submitted documents?	55%	32%	13%	59%	39%	2%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	60%	36%	5%	45%	49%	5%
To what extent is impact potential identifiable and measurable in the proposal?	43%	40%	17%	49%	43%	7%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	47%	37%	16%	46%	46%	7%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	47%	42%	12%	65%	31%	4%

Table A - 10. Stoplight assessment of data-collection and reporting credibility (by replenishment period)

Replenishment period	IRM (N=121)			GCF-1 (N=121)			GCF-2 (2024) (N=44)		
D. Data-collection and reporting credibility	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Are current reporting requirements sufficient for regular M&E?	28%	45%	26%	34%	50%	16%	61%	27%	11%
How feasible is it to measure and report credibly on the progress of investment criteria, given M&E plans, budget and indicators for investment criteria?	25%	59%	17%	46%	40%	14%	50%	43%	7%
To what extent did the proposal provide additional impact indicators beyond those proposed by the GCF? Can the proposal's indicators be used to measure the magnitude of causal change?	19%	21%	60%	38%	43%	19%	55%	36%	9%
Have baseline data been collected and/or is there a requirement for this?	27%	34%	39%	66%	28%	7%	43%	45%	11%
What is the potential quality of data, and are these suitable for IEs?	20%	31%	49%	34%	50%	16%	57%	32%	11%

Table A - 11. Stoplight assessment of data-collection and reporting credibility (by type of access)

ACCESS TYPE	IAE (N=216)			DAE (N=70)		
D. Data-collection and reporting credibility	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Are current reporting requirements sufficient for regular M&E?	37%	44%	18%	31%	44%	24%
How feasible is it to measure and report credibly on the progress of investment criteria, given M&E plans, budget and indicators for investment criteria?	40%	45%	15%	29%	59%	13%
To what extent did the proposal provide additional impact indicators beyond those proposed by the GCF? Can the proposal's indicators be used to measure the magnitude of causal change?	35%	30%	35%	24%	40%	36%
Has baseline data been collected and/or is there a requirement for this?	47%	31%	21%	42%	38%	20%
What is the potential quality of data, and are these suitable for IEs?	32%	39%	29%	30%	41%	29%

Table A - 12. Stoplight assessment of data-collection and reporting credibility (by changes of template: B.31 and before versus B.32 and after)

Replenishment period	B.31 and before (n=191)			B.32 and after (n=95)		
D. Data-collection and reporting credibility	% low risks	% medium risks	% high risks	% low risks	% medium risks	% high risks
Are current reporting requirements sufficient for regular M&E?	29%	48%	24%	51%	38%	12%
How feasible is it to measure and report credibly on the progress of investment criteria, given M&E plans, budget and indicators for investment criteria?	31%	52%	16%	50%	40%	10%
To what extent did the proposal provide additional impact indicators beyond those proposed by the GCF? Can the proposal's indicators be used to measure the magnitude of causal change?	24%	28%	48%	51%	41%	9%
Have baseline data been collected and/or is there a requirement for this?	40%	34%	26%	57%	31%	12%
What is the potential quality of data, and are these suitable for IEs?	24%	39%	38%	47%	41%	12%

Bibliography

Fiala, Nathan, Jyotsna Puri and Peter Mwandri. *Becoming Bigger, Better, Smarter: A Summary of the Evaluability of Green Climate Fund Proposals*. IEU Working Paper No. 1. Independent Evaluation Unit, Green Climate Fund, 2019. https://ieu.greenclimate.fund/sites/default/files/document/working-paper-becoming-bigger-better-smarter-summary-evaluability-gcf-proposals.pdf.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. *Impact Evaluation in Practice*. 2nd ed. World Bank, 2016. https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464807794.pdf?sequence=2&isAllowed=y.

Green Climate Fund. "GCF Project API". Accessed July 21, 2025. https://api.gcfund.org/v1/projects.

Fiala, Nathan, Saesol Kang, Peter Mwandri, Andreas Reumann, Shashank Sreedharan, and Galyna Uvarova. *Evaluability Assessment of the Green Climate Fund Funding Proposals*. Learning paper. Independent Evaluation Unit, Green Climate Fund, December 2022. https://ieu.greenclimate.fund/sites/default/files/document/230214-evaluability-study-top_2.pdf.

González-Caro, Alejandro, Peter Mwandri, Andreas Reumann, Galyna Uvarova, and Susumu Yoshida. *The Third Evaluability Assessment of the Green Climate Fund's Funding Proposals*. Independent Evaluation Unit, Green Climate Fund, May 2025. https://ieu.greenclimate.fund/sites/default/files/document/evaluability-study-iii-final-formatted-june-12th-1_0.pdf.

Independent Evaluation Unit. *Independent Evaluation of the Green Climate Fund's Energy Sector Portfolio and Approach*. Evaluation report No. 17. Independent Evaluation Unit, Green Climate Fund, February 2024. https://ieu.greenclimate.fund/evaluation/ES2023.

Jimenez, Emmanuel., and Jyotsna Puri. "The Wicked Cases of Education and Climate Change: The Promise and Challenge of Theory-Based Impact Evaluations". In *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*. Edited by Rob D. van den Berg, Indran Naidoo, and Susan D. Tamondong. International Development Evaluation Association (IDEAS), 2017. https://sdgs.un.org/sites/default/files/publications/2455IDEASwebREV_08Dec.pdf.

Organisation for Economic Co-operation and Development (2010). *Quality Standards for Development Evaluation*. DAC guidelines and reference series. OECD Publishing, 2010. https://www.oecd.org/en/publications/dac-quality-standards-for-development-evaluation_9789264083905-en.html.

———