IEU LEARNING PAPER

12/2022

Evaluability assessment of the Green Climate Fund funding proposals

Nathan Fiala, Saesol Kang, Peter Mwandri, Andreas Reumann, Shashank Sreedharan, Galyna Uvarova

GREEN CLIMATE FUND | Independent Evaluation Unit

Evaluability assessment of the Green Climate Fund funding proposals

Nathan Fiala, Saesol Kang, Peter Mwandri, Andreas Reumann, Shashank Sreedharan, Galyna Uvarova

12/2022

First Edition

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

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 information provided by the IEU is correct and verified. This document's author(s) or anyone connected with the IEU or the GCF cannot be held responsible for how its information is used.

About the IEU

The IEU was established by the Board of the Green Climate Fund (GCF) as an independent unit to provide objective assessments of the results of the Fund, including its funded activities, its effectiveness and its efficiency. The IEU fulfils this mandate through four main activities:

Evaluation: Undertakes independent evaluations at different levels to inform the GCF's strategic result areas and ensure its accountability.

Learning and communication: Ensures high quality evidence and recommendations from independent evaluations are synthesized and incorporated into the GCF's functioning and processes.

Advisory and capacity support: Advises the GCF Board and its stakeholders of lessons learned from evaluations and high quality evaluative evidence and provides guidance and capacity support to implementing entities of the GCF and their evaluation offices.

Engagement: Engages with independent evaluation offices of accredited entities and other GCF stakeholders.

About the IEU's Learning Paper series

The IEU engages with and contributes to debates on climate evaluation, low-emission pathways and climate-resilient development. The IEU Learning Paper Series encourages discussion on these topics and distributes findings from the Unit's work. The authors are IEU staff, consultants or are from partner organizations. Comments are welcome and should be sent to *ieu@gcfund.org.*

About this IEU Learning Paper

This paper continues IEU's effort to assess the tools and frameworks within funding proposals that enable robust monitoring and measurement of results during project implementation. The series began in 2019 with "Becoming bigger, better, smarter: A summary of the evaluability of Green Climate Fund proposals". This paper presents the progress in the ability of projects to deliver a cost-effective impact on climate adaptation and mitigation during the GCF-1 period.

ACKNOWLEDGEMENTS

LIST OF AUTHORS

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

FULL NAME	AFFILIATION
Nathan Fiala	University of Connecticut
Saesol Kang	Independent Evaluation Unit, Green Climate Fund
Peter Mwandri	Independent Evaluation Unit, Green Climate Fund
Andreas Reumann	Independent Evaluation Unit, Green Climate Fund
Shashank Sreedharan	New York University
Galyna Uvarova	Independent Evaluation Unit, Green Climate Fund

TABLE OF CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AE	Accredited entity
DAE	Direct Access Entity
EARF	Energy Access Relief Facility
FPs	Funding proposals
GCF	Green Climate Fund
IAE	International accredited entity
IE	Impact evaluation
IEU	Independent Evaluation Unit
LDC	Least developed country
M&E	Monitoring and evaluation
SIDS	Small island developing State
ToC	Theory of change

I. INTRODUCTION

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.

Following the first assessment conducted in 2019,[1] this document assesses the quality of proposals approved for financing by the GCF [also called funding proposals (FPs)]. It asks the following question: **To what extent are GCF supported programmes and projects capable of credibly reporting their impacts, efficiency and effectiveness in an evidence-based and robust way?**

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 GCF's contribution to the shift requires GCF investments to credibly commit to and measure the results to which they statedly aim. 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.[2] Results from individual actions on overall global climate change will only be apparent after hundreds of years, if not longer. However, we 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.

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.[3] 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.[4]

---

[1] Nathan Fiala, Jyotsna Puri and Peter Mwandri, "Becoming bigger, better, smarter: A summary of the evaluability of Green Climate Fund proposals", Working Paper No. 1 (Songdo, South Korea, Independent Evaluation Unit, Green Climate Fund, 2019). Available at *https://ieu.greenclimate.fund/sites/default/files/document/working-paper-becoming-bigger-better-smarter-summary-evaluability-gcf-proposals.pdf*.

[2] See Emmanuel Jimenez 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*, Rob D. van den Berg, Indran Naidoo and Susan D. Tamondong, eds., pp. 347-369 (Exeter, United Kingdom, International Development Evaluation Association (IDEAS), 2017). Available at *http://web.undp.org/evaluation/documents/Books/Evaluation_for_Agenda_2030.pdf*.

[3] GCF's investment framework (B.09/05).

[4] The overall criteria approved by the GCF Board for all IEU evaluations are: (1) relevance, effectiveness, efficiency, impact and sustainability of projects and programmes; (2) coherence in climate finance delivery with other multilateral entities; (3) gender equity; (4) country ownership of projects and programmes; (5) innovativeness in result areas, meaning the extent to which interventions may lead to paradigm shift towards low-emission and climate-resilient development pathways; (6) replication and scalability, meaning the extent to which the activities can be scaled up in other locations within the country or replicated in other countries (this criterion, which is considered in document GCF/B.05/03 in the context of measuring performance, could also be incorporated in independent evaluation); and (7) unexpected results, both positive and negative.

This study presents the results of an IEU desk assessment of the GCF portfolio. **The study has two main aims**. Firstly, to assess the quality of the proposals for the FPs the GCF has approved and is currently supporting. Project managers can learn from these and produce stronger proposals in the future that have a higher likelihood of reporting measured results and a greater likelihood of achieving success.[5] Secondly, 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.

These two aims help us meet **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 quality of proposals overall. Third, to help projects **measure** better and discuss methods FPs may use for this purpose. Hopefully, this discussion and the use of robust methods will enable the GCF to report its overall impact measurably and credibly.

This paper is organized as follows. Section II discusses the overall profile of GCF projects. Section III presents the methods used in this paper and discusses the development of a "stoplight" framework. Section IV presents our main findings, some common issues, standards for high quality measurement and possible learning tools. Last but not least, section V presents a comparative analysis across the GCF-1 versus IRM period funding proposals to assess how the quality at entry improves or changes over time.

---

[5] Our maintained hypothesis is that measuring impact (or the lack of it) will provide a good basis for improving the quality of subsequent investments since it will help the GCF learn faster.

II. SUMMARY OF THE 190 FUNDED PROJECTS APPROVED BY THE GCF

A. OVERVIEW OF GCF'S PORTFOLIO

Projects are the main route through which the GCF invests in low-emission, high-resilience development pathways. All GCF supported projects are expected to be "climate relevant": GCF investments are distributed among (i) projects that help developing countries reduce or mitigate their greenhouse gas emissions (mitigation projects), (ii) projects that help countries adapt to the increased likelihood of climate and weather shocks and increase the resilience of populations (adaptation projects), and (iii) projects that support both aims (cross-cutting projects). In this analysis, we examine proposals that have been approved for GCF investment – that is, proposals that are beyond stage 6 of the project's cycle.

Figure 1. GCF project/programme activity circle

Abbreviations: AE = accredited entity, NDA = national designated authority, ITAP = independent Technical Advisory Panel.
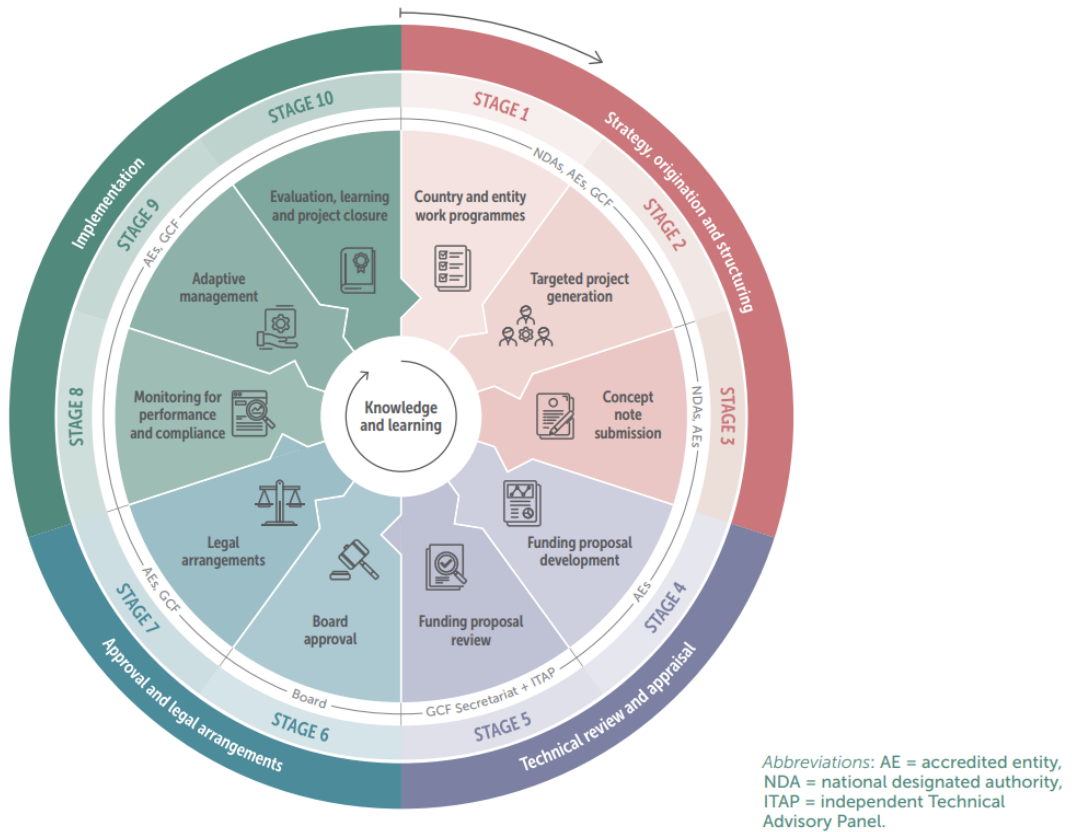
Source: GCF programming manual (2020).

This section presents a descriptive summary of 190 approved GCF proposals. Submitted by 127 countries, some 41 per cent of total proposals were made by least developed countries (LDCs) and 24 per cent by small island developing States (SIDS). Around one-fifth of the proposals were multi-country efforts. The smallest multi-country proposal comprises two participating states, while the largest comprises 42. The Africa and Asia-Pacific regions (see Table 1) are the most represented GCF regions, accounting for 82 per cent of all approved proposals. In contrast, the Latin American Caribbean, and Eastern European regions, account for 18 per cent. Thematically, adaptation proposals constitute nearly half of all approved proposals (43 per cent), while mitigation and cross-

cutting proposals account for around 32 per cent and 25 per cent, respectively (Table 2). The public sector is the most represented, with 79 per cent of approved proposals.

This range of successful proposals was spoken for by 48 different accrediting entities, most of them direct access entities (DAEs) (25). However, they managed to channel only 19 per cent of the approved finance compared to 23 international accredited entities (IAEs) with 81 per cent of all approved finance (USD 9.9 billion). The average GCF financing awarded per country is USD 78.03 million, while the median is USD 51.9 million.

Table 1.   Regional distribution of the GCF portfolio

GCF REGION	NUMBER OF PROPOSALS	PER CENT (%)	VOLUME OF FINANCE (GCF), USD	PER CENT (%)
Africa	77	36	3,592.8m	36
Asia-Pacific	79	37	3,469.6m	35
Eastern Europe	11	5	393.9m	4
Latin America and the Caribbean	46	22	2,462.3m	25
Total	190*	100	9,918.5m	100

Source: GCF Tableau server as of 31 December 2021.
Note: * The number of proposals does not aggregate into a total due to some approved proposals covering several regions.

Table 2.   GCF portfolio by thematic distribution

THEME	NUMBER OF PROPOSALS	PER CENT (%)	VOLUME OF FINANCE (GCF), USD	PER CENT (%)
Adaptation	82	43	2,389.5m	24
Cross-cutting	48	25	3,046.2m	31
Mitigation	60	32	4,482.8m	45
Total	190	100	9,918.5m	100

Source: GCF Tableau server as of 31 December 2021.

B. BREAKDOWN OF RESULTS

Figure 2 below shows the distribution of approved proposals by GCF region. However, we are also interested in more than a bird's eye view of this breakdown. First, looking at the distribution of GCF regions by LDC, we find that the most successful LDC proposals were from Africa. In contrast, far fewer proposals for other regions came from LDC-identifying countries, including none from Eastern Europe. Similarly, the Asia-Pacific and the Latin America-Caribbean regions account for most SIDS, as we might expect.

Africa also accounts for the most multi-country proposals (46 per cent), followed by Latin America and the Caribbean, while the Asia-Pacific accounts for the most single-country proposals (43 per cent).

Thematically, most successful proposals from Africa were aimed at adaptation (44 per cent), as was the case for the Asia-Pacific region.[6] However, most proposals from Latin America, the Caribbean and Eastern European aimed at mitigation.

---

[6] It is important to note that multi-country projects approved for more than one region result in double counting of the number of proposals.

Next, we found that the public sector was the most represented among the approved proposals. For example, of the 77 successful proposals from the Africa region, 53 are public sector proposals against 24 private sector ones. Eastern Europe, Latin America and the Caribbean saw about twice as many public than private sector proposals. Moreover, we found four times more public sector proposals from the Asia-Pacific than private sector proposals. Finally, the average GCF financing per region is highest for the Asia-Pacific.

Figure 2.   Approved funded projects by region
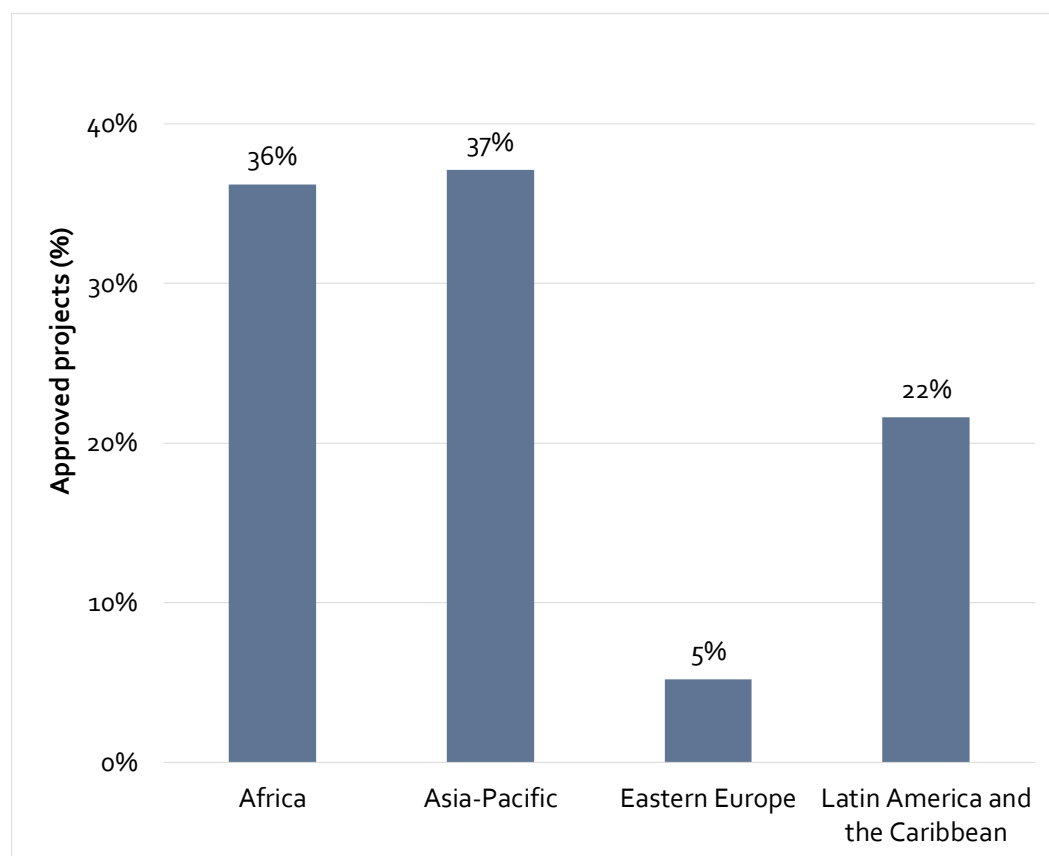
Source: GCF Tableau server as of 31 December 2021.

Table 3.   Distribution of the GCF portfolio by region and vulnerable group

	LDCs			SIDS		
GCF region	Number of FPs	Share (%)	Volume of finance (USD M)	Number of FPs	Share (%)	Volume of finance (USD M)
Africa	53	28%	1,938.9	8	4%	194.0
Asia-Pacific	23	12%	932.2	21	11%	571.4
Latin America and the Caribbean	1	1%	13.9	16	8%	410.2
Total	77	41%	2,885.0	45	24%	1,175.6

Source: GCF Tableau server as of 31 December 2021.

C. PORTFOLIO BY LDCs AND SIDS

Two-fifths of successful proposals came from LDCs, while SIDS represent 24 per cent of proposals approved for GCF financing. Sixteen proposals were submitted by countries that are both LDCs and SIDS.[7]

Among the successful LDC proposals, 38 per cent were multi-country efforts. In terms of theme, around 44 per cent of all successful LDC proposals were aimed at adaptation, whereas for SIDS, adaptation proposals account for more than half of the GCF financing (see Figure 3). In terms of sector, more than 80 per cent of LDC and SIDS proposals are public sector.

Considering the most vulnerable groups of countries, the average levels of GCF financing per country for LDC and SIDS proposals were USD 68.7 million and USD 30.9 million, respectively. For the 16 proposals that were both LDC and SIDS, the average financing per country was USD 20.1 million.

Figure 3.   GCF portfolio in SIDS and LDCs
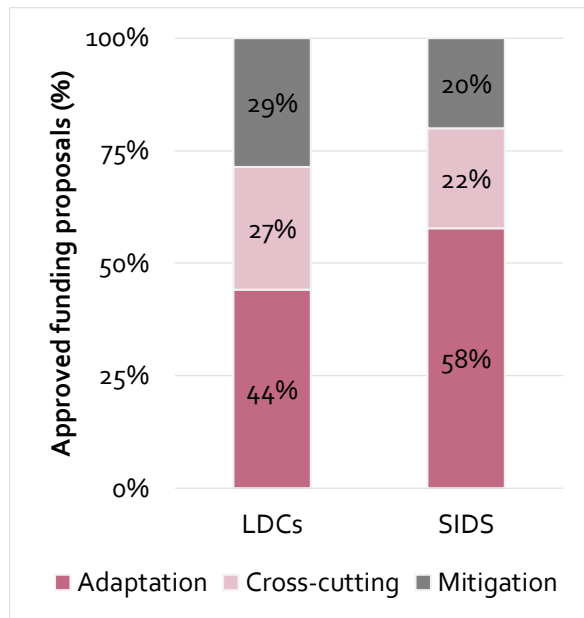
Source: GCF Tableau server as of 31 December 2021.

D. PORTFOLIO BY COUNTRY FOCUS

Around one-fifth of all successful proposals are multi-country. Among these, the majority (40 per cent) are mitigation projects, while adaptation and cross-cutting projects each have a 30 per cent share of successful proposals. Furthermore, most multi-country projects are private sector projects. However, the average GCF financing per country is higher for single-country projects (USD 69.9 million) than for multi-country projects (USD 37.9 million).

---

[7] Please note that some countries can exist in more than one category. For instance, Tuvalu appears under LDC and SIDS.

Figure 4.   GCF portfolio by country focus
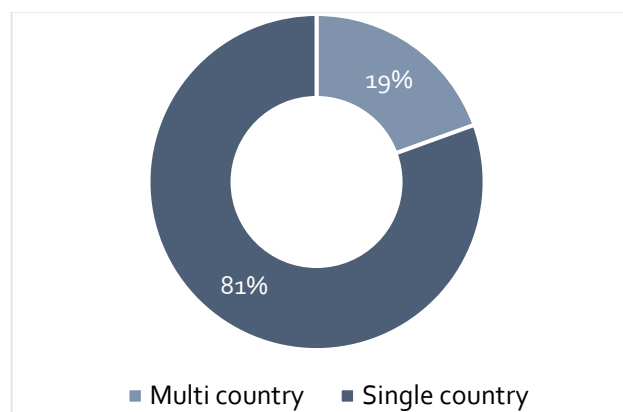
Source: GCF Tableau server as of 31 December 2021.

Table 4.   Distribution of GCF portfolio by country focus

	CATEGORIZATION	MULTI-COUNTRY PROJECTS	SINGLE-COUNTRY PROJECTS	TOTAL
THEMATIC FOCUS	Adaptation	11	71	*82*
	Cross-cutting	11	37	*48*
	Mitigation	15	45	*60*
SECTOR	Private	20	19	*39*
	Public	17	134	*151*
TOTAL		*37*	*153*	190

Source: GCF Tableau server as of 31 December 2021.

E. PORTFOLIO BY THEME AND SECTOR

A closer look at the thematic breakdown of the proposals awarded GCF financing reveals nearly half of all successful proposals focused on adaptation (43 per cent). In comparison, about one-third focused on mitigation (Figure 5). A significant amount of adaptation proposals were public sector projects (93 per cent), with only five pure adaptation projects from the private sector. The private sector portfolio is predominantly mitigation projects (87 per cent) when including cross-cutting projects.

Figure 5.   GCF portfolio by thematic focus
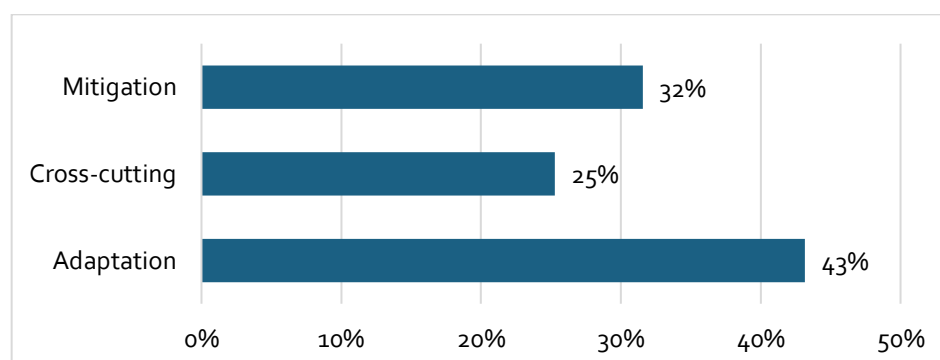
Source: GCF Tableau server as of 31 December 2021.

Mitigation proposals had the highest relative GCF financing per country, with an average of USD 51.09 million, followed by cross-cutting proposals, with an average of USD 39.4 million per country. In terms of sector, public sector projects have higher average financing per country of USD 58.8 million relative to private sector projects, which have average financing of USD 42.2 million per country.

Table 5.   Distribution of projects by sector

THEME	PRIVATE	PUBLIC	TOTAL
Adaptation	5	77	82
Cross-cutting	9	39	48
Mitigation	25	35	60
Total	39	151	190

Source: GCF Tableau server as of 31 December 2021.

III.   METHODS OVERVIEW

This section discusses the methods used to assess proposals submitted and approved for funding by the GCF. We ask, "**To what extent are approved projects likely to be able to inform results that they claim in a credible and measured manner?**"

Board-approved FP proposals were assessed along several dimensions. These dimensions are grouped into several common topics.[8] An important caveat here is that the comments in this study are not indicative of the capacity or ability of proposal submitters, mainly because – to the extent we know – the attributes we analyse were neither GCF requirements nor funding conditions when these proposals were prepared. Inferences made in this paper are made from data and information in the proposals. They indicate whether proposals and the information within them are fit for purpose in fulfilling the GCF's objectives. While noting this is a desk review, our guiding principle is only to examine what has been submitted in proposals. We also provide constructive comments on how these proposals may be improved and comment mainly on the potential "internal validity" of these projects – that is, the feasibility and ability of the projects to inform the results/changes they aspire to, as stated in their proposals.

A.   LENS TO ASSESS WHETHER PROPOSALS ARE FIT FOR PURPOSE

We use four lenses to assess the potential for the internal validity of FP proposals. The first is the **theory of change** (ToC). We assess whether proposals include an explicit discussion of the project's overall ToC. A clear ToC is critical for understanding if the proposed activities will lead to the investment's intended outcomes, including their size. In many FP proposals, ToCs are not laid out coherently in a single place. In all cases, we analyse the full proposal and piece together an "implicit" ToC if an explicit one is omitted.

Second, to understand if the programme activities can achieve the impacts claimed in the proposal, we examine proposals for their potential to **measure** and report **causal change and report results using impact measurement**. Many proposals make bold claims about what their investments can accomplish. This causality and impact lens enables us to objectively estimate if the project investment will cause the claimed impact(s) or whether some proportion of the anticipated effect would still have occurred in the absence of programme activities/GCF investment. Observing this counterfactual scenario is impossible: we cannot observe what happens to a beneficiary if it receives a project intervention and what happens if it does not. But there is now a sizeable discipline showing this can be done using experimental construction or observation of valid comparison groups to generate accurate estimates of causal impacts.[9]

Third, we assess the project's ability to inform the GCF investment criteria credibly. We assess the extent to which the proposal's credibly responds to fulfilling the GCF's required investment criteria, including whether the proposal demonstrates a strong potential for delivering impact, paradigm shift, sustainable development, recipient needs, country ownership, effectiveness and efficiency.

---

[8] The individual assessment documents present our detailed comments on each of the proposals and are available on request.

[9] See, for example,

(1) Paul J. Gertler, and others, *Impact Evaluation in Practice*, 2nd ed. (Washington, D.C., World Bank, 2016). Available at *https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464807794.pdf?sequence=2&isAllowed=y*;

(2) Emmanuel Jimenez 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*, Rob D. van den Berg, Indran Naidoo and Susan D. Tamondong, eds., pp. 347-369 (Exeter, United Kingdom, International Development Evaluation Association (IDEAS), 2017). Available at *http://web.undp.org/evaluation/documents/Books/Evaluation_for_Agenda_2030.pdf*.

Specifically, we recognize that **targeting** is a primary concern for many projects. We assess each project based on its targeting criteria. For instance, if a programme plans to reach 50 per cent of women or vulnerable groups, we examine if the proposal has articulated its targeting criteria clearly, and the extent to which the programme is likely to achieve this goal, based on the programme model and ToC.

Fourth, we examine FPs for how well they have set up systems to help report on their progress and **fidelity to implementation** plans. We also examine their stated monitoring and evaluation (M&E) systems to see if they are sufficient in their current state to assess the projects' capabilities in this area.

B. BUILDING A STOPLIGHT

To illustrate results, we built a **stoplight** for each FP proposal that summarizes risks and other issues related to results measurement and the information presented in each FP. Four criteria inform the stoplight. We use a likelihood or risk framework for each criterion to assess the quality with which the proposal meets each requirement (see Box 1). The following decision rule is used:

- If the FP has done well on a criterion, and it is highly likely that the criterion will be achieved, the proposal is marked as "low risk" for that criterion.

- If, based on the information provided in the FP, there appears to be a moderate probability the proposed programme or project will perform well relative to the stoplight criterion, then the proposal is marked as "medium risk" for that criterion.

A proposal is marked as "high risk" for a given criterion if there appears to be a high probability that the proposed programme or project will not perform well relative to the criterion.

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.

C. THEORY OF CHANGE AND DISCUSSION OF CAUSAL PATHWAYS

We use the following questions and rating rules to assess the quality of the ToCs 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 for a ToC checklist.)**

   a) Low risk. ToC is well articulated.

   b) Medium risk. Logic framework or ToC is present but needs some clarification. (Missing information is specified.)

   c) High risk. Logic framework or ToC 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 ToC.

2) **Are unintended consequences referred to and identified robustly in the programme ToC and/or in the surrounding literature reviews?**

   a) Low risk. Unintended consequences are well articulated. (These are drawn from discussion of the ToC.)

   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 ToC 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.

D. POTENTIAL FOR MEASUREMENT OF CAUSAL CHANGE AND EVALUABILITY

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.

E. IMPLEMENTATION FIDELITY AND PERFORMANCE AGAINST INVESTMENT CRITERIA

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.

F. DATA COLLECTION AND REPORTING CREDIBILITY

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 evaluate 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 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.

We use these questions to assign each FP to high risk, medium risk, low risk or unclear categories for each of the questions. Assigning FPs to different risk categories should be made in conjunction with the individual project assessments (available on request). This will help the reader better understand the impact and feasibility of the proposals. Each approved project is assessed using these four lenses. The table summarizes the basic information of the proposal (number, implementer, period of funding, countries and funding amount) and the results of the assessments, focusing on the ToC's quality, causal linkages, targeting strategy, and whether the proposal can rigorously inform the GCF investment criteria. We then aggregate these rankings. The overall results are discussed in the next section.

.

IV. STOPLIGHT ANALYSIS

This section looks closely at the stoplight data for any meaningful trends and inferences from the proposal evaluations. We summarize overarching trends and assess differences by key factors such as theme, country focus, and sector.

The stoplight data is used to assess the quality of the proposal across four categories:

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

In category A, we assess the extent to which the pathways to impact are outlined using a ToC or logic model. The assessment seeks to determine if the ToC 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 and used to build their programme (Figure 6).

We find that most proposals, explicitly or implicitly, outline their programme logic and reasonably substantiate the credibility of their claims about causal pathways, with results skewed towards medium risk. Some 36 per cent of approved FPs cite good evidence supporting their causal claims. However, an area for improvement is acknowledging and planning for unintended consequences; 34 per cent of proposals satisfactorily account for the unintended consequences of their GCF funding, with 28 per cent leaving the matter unaddressed.

Figure 6.   Stoplight assessment of the theory of change
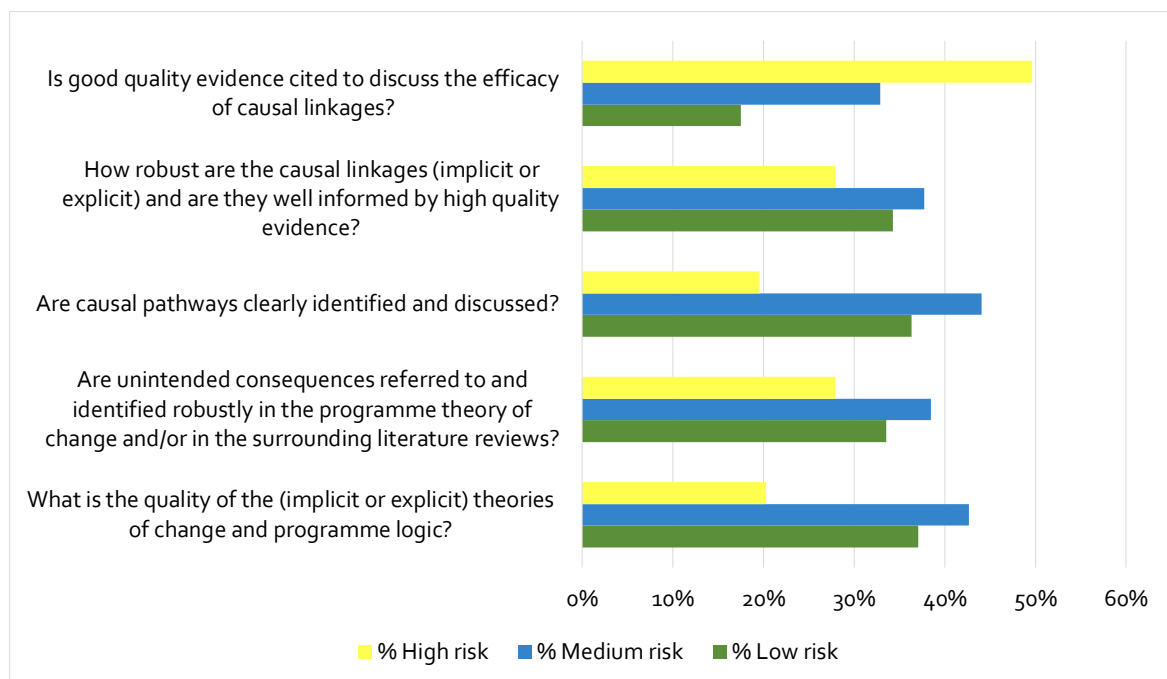
Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

In category B, we assess the evaluability and feasibility of measurement of the proposed project activities. In other words, we seek to determine if the (claimed) causal effects of the proposed activities can be credibly measured by examining the FP's plans for M&E.

While most FPs refer to planned M&E activities or budgetary allocation, very few provide specific information regarding their evaluation strategies and planned actions to ensure attributable change is measured. Several FPs claim they will track results over time. Still, these are either unaccompanied by any qualifying statements on their methodological choices or neglect the temporal and spatial risks to their estimates. As seen in Figure 7, most of the FPs are assessed as having medium and high risk, particularly in the context of methods for measuring causal change and safeguards against biased estimates of impact.

Therefore, a key trend worth pointing out in this report is an apparent misunderstanding regarding the role of M&E in measuring causal change. Monitoring and tracking changes in key outcomes over time (progress tracking) are inadequate for causal measurement. This approach cannot detect effects due solely to the programme that *would have been missing* if the programme did not exist (known as a counterfactual or comparison group).

Figure 7.   Stoplight assessment of the potential for measurement of causal change and evaluability
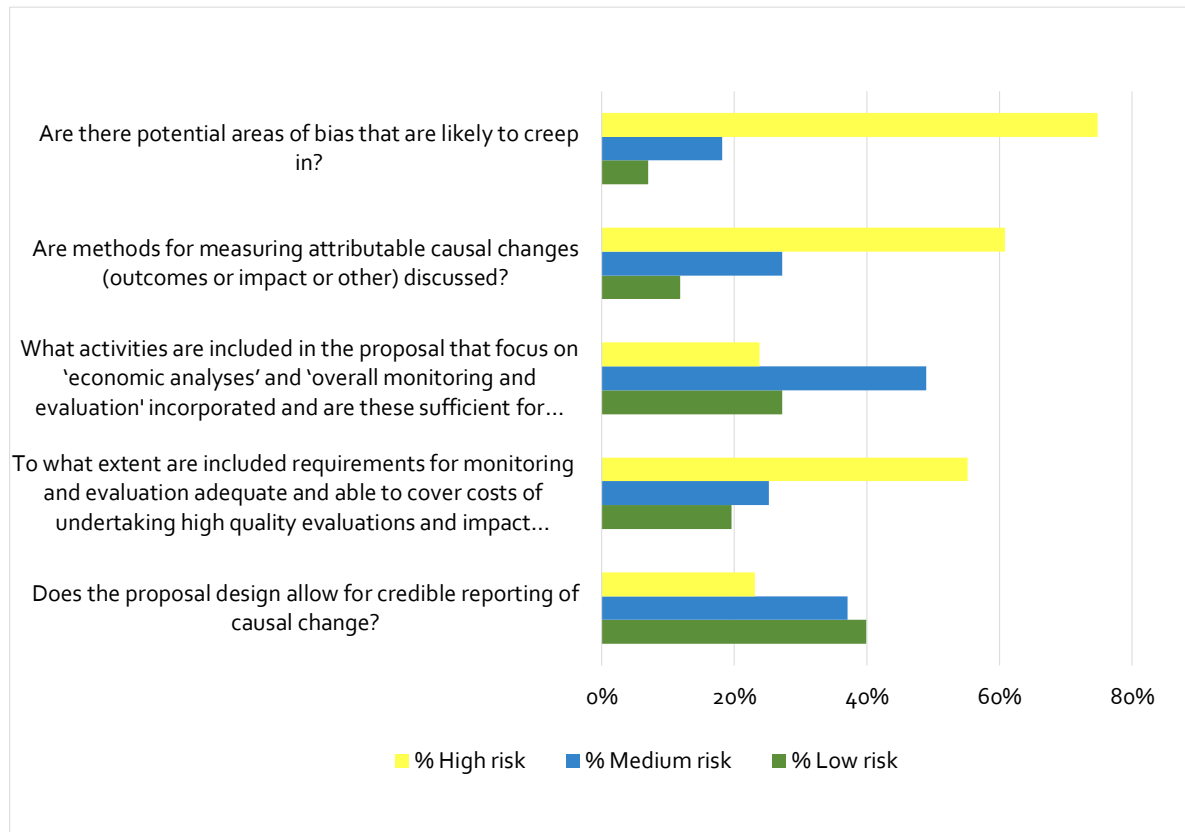
Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

In category C, we seek to assess if the project activities are well-targeted and to determine the programme's performance against GCF's investment criteria, the feasibility of the implementation plans for the overall programme, and if the proposal identifies relevant barriers to implementation and includes plans for recourse in the event of such constraints.

As seen in Figure 8, most FPs can be determined as a low or medium risk since their description of their implementation plans and their performance against GCF's investment criteria are thorough. However, some FPs do not fully establish the links between their performance against investment criteria and their specific activities and outputs in their ToC. Additionally, 63 per cent of all FPs are careful to identify potential risks to implementation fidelity and discuss steps to address them.

Figure 8. *Stoplight assessment of implementation fidelity and performance against investment criteria*
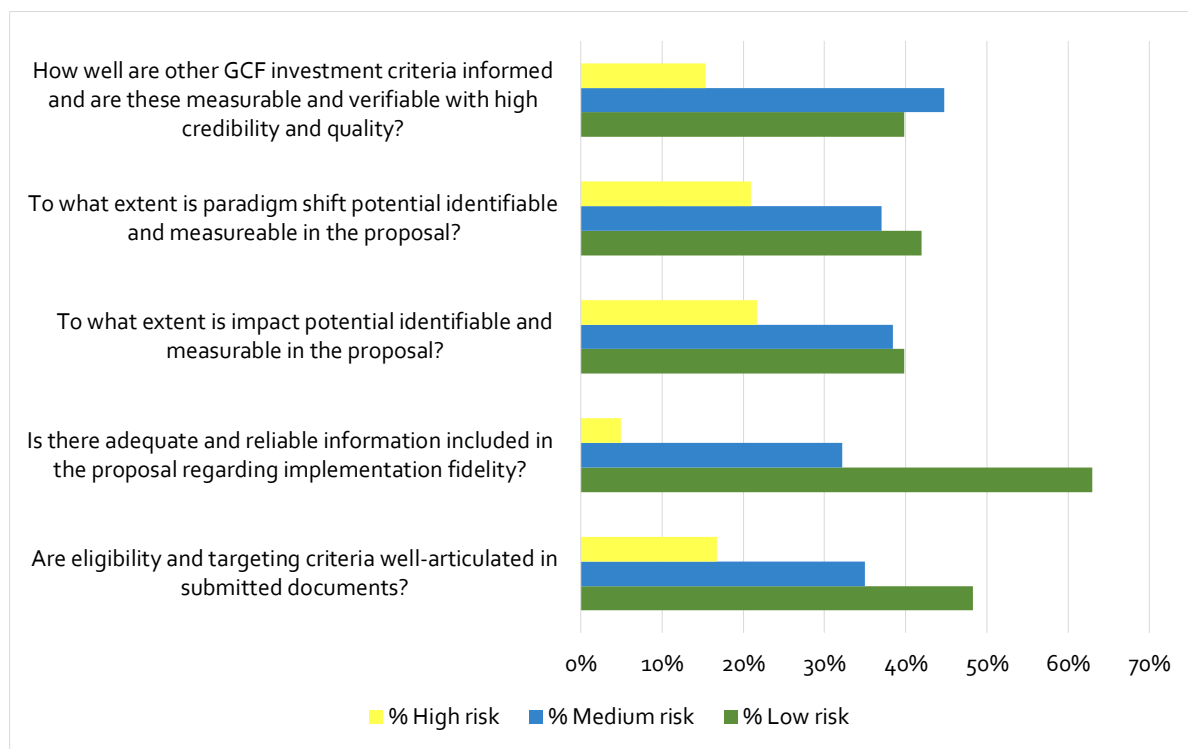
Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

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 (Figure 9). First, 36 per cent of FPs indicated they were going to collect or already had adequate baseline data for evaluative purposes, putting them at low risk. Another 33 per cent referred to plans of developing a baseline or equivalent exercise (such as acquiring relevant administrative records). Furthermore, only 27 per cent of FPs are thorough in their description of the frequency and level of data collection and reporting necessary to ensure M&E activities carry on unhindered. While the data collection and reporting requirements are deemed mainly sufficient for overall M&E, an area needing improvement is the level of transparency and preparedness in the context of a full-fledged evaluation. However, over 80 per cent (low and medium risk) of FPs have systems and budgetary allocations either in place or in their plans to conduct the evaluation design in their FP. While the evaluation plans might need revisiting, their data collection and reporting plans follow their proposed evaluation strategies.

Figure 9. Stoplight assessment of data collection and reporting credibility
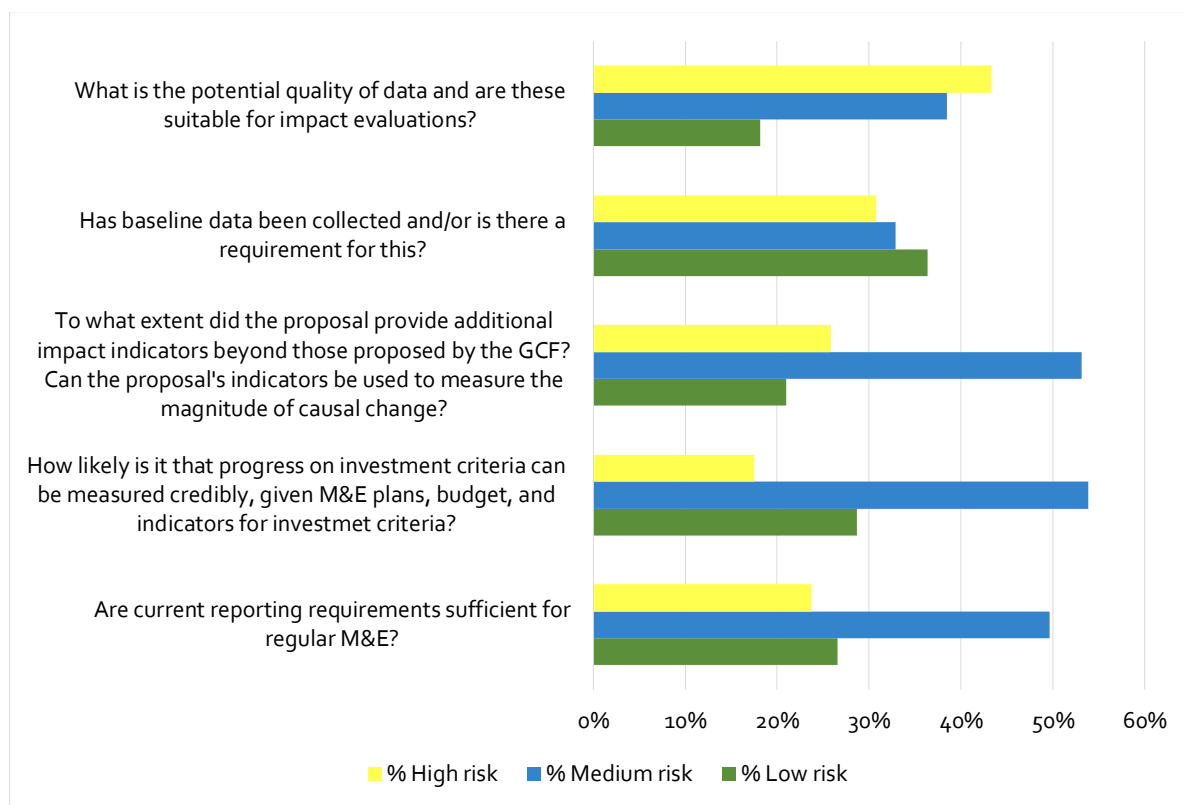
Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

A. STOPLIGHT ASSESSMENT BY THEME

This section looks at stoplight results and trends disaggregated by FP theme. First, we find that most adaptation FPs are assessed as low or medium risk, skewed towards low in categories A, C, and E and towards medium in category B. A persistent trend is the lack of attention to identifying unintended consequences, with 23 per cent of all adaptation proposals deemed high risk. The main areas of concern to emerge from the adaptation FPs occur in category B—specifically, making explicit methods to measure both causal impact and safeguarding against biased estimates, wherein 58 per cent and 75 per cent of proposals are marked as high risk in these categories, respectively. On the flip side, adaptation FPs systematically discuss performance against GCF investment criteria.

Unlike our experience with adaptation proposals, we were able to determine that the mitigation FPs were more medium and high risk regarding their ToCs' strength and their discussion of causal change measurement and evaluability. Notably, we see persistent trends concerning the lack of attention given to identifying and discussing unintended consequences, with 80 per cent of proposals being marked as medium or high risk, with the latter reaching 40 per cent. Moreover, 52 per cent of mitigation proposals were determined as medium risk in their discussion of causal pathways. Similarly, 42 per cent did not adequately cite rigorous evidence to support their causal linkages and were also marked medium risk. Given the activity-heavy programmes associated with mitigation projects, it is interesting to note that mitigation FPs are heavily skewed towards medium rather than low risk in category C. An indicator that stands out is implementation fidelity. We determine nearly 50 per cent of mitigation proposals to be medium risk, while 61 per cent and 67 per cent of proposals, respectively, do not sufficiently provide for measuring paradigm shift potential and measuring against GCF investment criteria. Given the assessments for category B, it is not

unexpected that their data collection and reporting sections are also determined as, on average, being of medium risk.

Finally, we see a much more even spread of FP quality for cross-cutting projects. Regarding programme theory models, we see an even distribution of well articulated projects with strong ToCs, FPs with reasonably enumerated ToCs but underwhelming supporting evidence, and ToCs that do not clearly articulate their pathways to impact or why they are cross-cutting. In terms of category B, however, they are primarily determined to be low or medium risk, except for sizeable concerns around measuring impact (63 per cent high risk) and handling bias (71 per cent high risk). This further strengthens our stoplight's finding that FPs refer to M&E plans but without exploring methods to measure attributable change. However, in categories C and D, cross-cutting FPs are more articulate in describing their implementation plans, data collection and reporting strategies, and identifying risks to targeting and implementation fidelity.

B. STOPLIGHT ASSESSMENT BY LDC/SIDS

Aside from disaggregating stoplight results by theme, we are specifically interested in seeing if any trends emerge in proposals submitted by United Nations priority countries designated as LDCs and SIDS. This section discusses the differences between LDC (SIDS) and non-LDC (non-SIDS) FPs.

In category A, there were no discernible differences in FPs submitted by LDCs. While they had a marginally higher number of high risk proposals on identifying causal pathways, they fared considerably better in articulating the causal linkages in their ToCs relative to non-LDC FPs. In terms of category B, however, LDC FPs appear to be more likely to be medium and high risk, particularly in being prepared to measure causal change (26 per cent medium risk and 65 per cent high risk), and deal with potential bias in their measurement and estimates of impact (21 per cent medium risk and 74 per cent high risk). Similarly, in category C, a slightly higher proportion of LDC FPs were adjudged as low risk regarding indicators for identifying and measuring paradigm shift potential (12 per cent for LDC FPs versus 4 per cent for non-LDC FPs), and performance against GCF investment criteria (44 per cent versus 35 per cent). Finally, in category D, both LDC and non-LDC FPs fared similarly. Still, the non-LDC turned in a higher number of high risk proposals regarding measuring progress against GCF investment criteria (26 per cent). In terms of providing additional relevant indicators to measure progress, approximately 23 per cent of accepted proposals on both sides of the LDC aisle were marked high risk. These are consistent with our assessments in category C.

These inferences indicate that LDC's submitted FPs might need more assistance in better articulating their strategies, looking ahead, and reasoning back to identify potential barriers to successful implementation.

We observe similar differences with the SIDS comparison. SIDS FPs appear to have a consistently higher number of high risk ToCs than non-SIDS proposals and a much lower rate of low risk submissions. Some 38 per cent and 35 per cent of SIDS-submitted proposals were determined to be high risk in their identification and discussion of unintended consequences and the robustness of the causal linkages in their ToC. In category B, however, while non-SIDS FPs are evenly distributed in terms of risk, SIDS-submitted proposals tended to be substantially centred around medium risk. We see 79 per cent of SIDS proposals marked as high risk on the likelihood of potential bias. A key area of concern for SIDS proposals is their lack of focus on measuring causal effects (29 per cent high risk) and lack of discussion acknowledging the enormous risk of bias in their evaluation strategy. In terms of category C, SIDS FPs fared similarly to non-SIDS ones. They even did better on indicators like identifying paradigm shift potential and measuring performance against other GCF investment criteria, as evidenced by not having any high risk submissions against those indicators. Finally, we

see that SIDS-submitted proposals appear to be more skewed towards medium risk relative to non-SIDS submissions in category D. SIDS proposals also tend to have more high risk marked proposals, with 26 per cent of their submissions deemed high risk in terms of their current M&E reporting requirements, and 35 per cent estimated as high risk for their provision of additional impact and outcome indicators.

C. STOPLIGHT ASSESSMENT BY COUNTRY FOCUS

As shown in the overall summary of results section above, one-fifth of all submitted proposals were multi-country efforts. As a result, we examine the differences between the risk assessments of multi-country and single-country proposals. Starting with category A, we observe that more high risk ToCs were submitted in single-country FPs. In contrast, multi-country FPs significantly neglected to acknowledge and plan for the unintended consequences of their programmes in their submissions. In fact, 52 per cent of multi-country accepted proposals were marked high risk in their identification of externalities. Single-country proposals skew to medium risk as well, but they fare much better in identifying unintended consequences at 22 per cent high risk, while 33 per cent and 50 per cent of single-country submissions are low-risk regarding their robustness in identifying causal linkages and using rigorous evidence in support of their causal claims.

Category B sheds light on an interesting split. On average, multi-country FPs are much more high risk regarding discussing plans to measure causal impact and in their overall M&E plans and resource allocations. Some 38 per cent of multi-country proposals are high risk in their planning for credible reporting of causal change and safeguarding against bias in their estimates. Of more concern is that 55 per cent of accepted multi-country proposals are deemed high risk in their choice of methods for measuring attributable change. Although we can only speculate, this might be because of differing M&E standards when participating states are various and diverse. However, this is linked to a more significant issue of FPs – and eventually, projects/programme – without the capacity to evaluate the impact of their programmes. Potentially it can be resolved using an empirically informed technical standard for conducting IEs instead of simple progress monitoring. Overall, multi-country FPs were consistently more high risk, whereas single-country submissions were skewed towards low and medium risk.

Both kinds of submissions appear more evenly distributed in category C. There are fewer low risk and more high-risk multi-country submissions than single-country, with the identification and measurement of impact potential of particular concern for multi-country proposals (34 per cent high risk). This is consistent with our inferences from category B. Finally, single-country FPs are largely low or medium risk in category D, while multi-country FPs skew towards medium and high risk. Like our inferences above, multi-country FPs do not adequately discuss measuring progress against GCF investment criteria or contributing relevant indicators beyond GCF's pre-requisite list of outcome indicators. This reveals the overall lack of strength in their plans for measuring the impact of their programmes in multi-country proposals.

D. STOPLIGHT ASSESSMENT BY SECTOR

In addition to country focus, we are also interested in disaggregating our stoplight data by whether the accrediting agency belongs to the private or public sector. Starting with the ToC, we find that public sector entity-supported FPs are much more skewed towards low and medium risk. In contrast, private sector supported entities are more skewed towards medium and high risk. In particular, submissions of the latter kind generally fail to consider the unintended consequences of the GCF investment (41 per cent high risk). Further, they do not cite sufficiently robust and scientific

evidence in support of their causal claims in the ToC (31 per cent high risk and 28 per cent medium risk).

We see a similar but less pronounced trend in looking at the respective causal change and evaluability strategies. While more public sector entity-supported FPs appear low risk, both types of submissions are faced with high risk submission elements. The most concerning of these is the lack of attention to measuring causal effects and the accompanying lack of preparedness to deal with the potential bias of any measurement efforts. Private sector proposals were marked at 53 per cent and 84 per cent high risk respectively on these indicators, while their public sector counterparts were marked at 63 per cent and 72 per cent high risk. Furthermore, 23 per cent of public sector proposals were also estimated as high risk on their inclusion of economic analysis and overall M&E. Also, nearly 30 per cent of private sector proposals were high risk on the remaining category B indicators pertaining to credible reporting of causal change and their M&E preparedness.

Public sector entity-supported FPs fare much better in category C, with their submissions substantially low or medium risk. While private sector-supported proposals are also mainly medium risk, the high risk submissions cannot be ignored, especially the measurement of impact potential (19 per cent high risk) and the articulation of eligibility and targeting criteria (25 per cent).

Finally, we observe a set of differences in terms of the data collection and reporting requirements similar to category C. Public sector entity-supported submissions skew towards low or medium risk. In contrast, private entity-supported FPs are more evenly distributed. The higher risk elements are the quality of data conditional on conducting IEs (44 per cent high risk) and their capacity to measure progress against GCF investment criteria (22 per cent high risk), especially without including any additional outcome indicators of relevance (31 per cent high risk).

E. STOPLIGHT ASSESSMENT BY ACCREDITED ENTITY MODALITY

We are interested in examining trends in the stoplight data by the modality of the accredited entity (AE). Starting with category A, we find that the risk levels in DAE proposals appear to be evenly distributed. Regarding ToC quality, each risk category hosts nearly a third of all accepted DAE proposals. Some 23 per cent of DAE proposals are marked medium risk for identifying unintended consequences. Two areas of concern are the enumeration of the causal pathways and the evidence used to support them. With both these indicators, 35 per cent and 42 per cent of DAE proposals are estimated as high risk, respectively. On the other hand, IAE proposals appear at face value to skew towards low and medium risk. Across indicators in category A, around 35 per cent of successful IAE proposals are low risk. The evidence used to substantiate the causal pathways in the ToC has improved, since nearly 54 per cent of proposals are either medium or high risk. Similarly, approximately 71 per cent of IAE proposals are marked medium or low risk in their discussion of unintended consequences.

The risk assessment for category B tends to fluctuate. Among DAE proposals, more than half (55 per cent) are marked low risk for credible reporting of causal change. However, the M&E requirements in general, and the methods for measuring attributable change in particular, are considered high risk. The current M&E reporting requirements are largely underwhelming, with 55 per cent of proposals being high risk. The discussion on methods to measure causal change is equally weak, with 58 per cent of proposals being high risk. Perhaps most significant is the assessment that 87 per cent of all successful DAE proposals are at high risk of having biased estimates of impact. The critical inference underlying these trends is a general lack of rigour in the proposal's M&E and IE strategies.

In contrast, IAE proposals appear to be more high or medium risk. Some 51 per cent of proposals are low risk in their use of quality evidence to support causal claims, while 30–37 per cent of all

IAE proposals are low risk against other category A indicators. The robustness of the causal linkages and the identification of unintended consequences are the major areas of improvement, with 24 per cent and 29 per cent of proposals identified as being high risk in these areas, respectively.

DAE proposals are predominantly low risk in category C. Notable exceptions include the measurement of paradigm shift potential, wherein 23 per cent of proposals are marked high risk, and an additional 35 per cent are assigned medium risk. Moreover, 52 per cent of proposals are assessed as medium risk for measuring impact potential. In comparison, 39 per cent of accepted DAE proposals are estimated as medium risk when measuring progress against GCF investment criteria. Notably, nearly 68 per cent of DAE proposals are low risk in their inclusion of economic analysis and comprehensive M&E plans.

On the other hand, IAE proposals appear to be largely low or medium risk and show some distinct areas of improvement. The percentage of high risk proposals is relatively higher on measurement indicators, namely for impact potential (23 per cent), paradigm shift potential (21 per cent), and progress against GCF investment criteria (17 per cent). IAE proposals also appear to perform well in their inclusion of economic analysis and M&E plans (62 per cent low risk). Approximately 35–46 per cent of all IAE proposals are estimated as medium risk across all category C indicators.

In category D, accepted DAE proposals are heavily skewed towards medium risk. Two significant sources of risk are plans for baseline data collection and the quality of IE data. For these indicators, 32 per cent and 42 per cent of all DAE proposals are marked high risk, respectively. This is somewhat contrary to proposals including economic analyses and M&E plans, suggesting that while DAE proposals generally outline their M&E strategies, the focus on IE and measuring casual change is missing.

IAE proposals, while also medium risk-skewed, are more evenly distributed regarding risk assessment. The quality of potential data collected is a significant source of high risk at 44 per cent, while 30 per cent of IAE proposals are demarcated as high risk due to their baseline data collection plans (or lack thereof). Some 28 per cent of IAE proposals are determined to be high risk for not providing any additional impact indicators, while 23 per cent are considered high risk due to a lack of confidence in their current M&E reporting processes.

F. DISCUSSION

The assessment observed that most (80%) funding proposals are good in explicitly or implicitly outlining their programme logic model and reasonably substantiating the credibility of their claims about causal pathways, with results skewed towards medium-to-high risk. However, some FPs do not cite good evidence to support their causal claims and need to improve acknowledging and planning for unintended consequences in their programme's ToC.

A funded project titled "Scaling Smart, Solar, Energy Access Microgrids" in Haiti aims to build an additional 22 community-scale solar photovoltaic microgrids in southern Haiti in communities without grid power. The project will provide affordable and reliable access to modern energy services in communities previously identified through extensive market scoping. The project presented the ToC in both FP and pre-feasibility study annexes and focused on project activities. However, the proposal lacks any discussion of the unintended consequences of the project and the ToC's causal pathways.

In the case of the funded project "Building climate resilience of vulnerable and food insecure communities through capacity strengthening and livelihood diversification in mountainous regions of Tajikistan", two components are worth noting: (i) Capacity strengthening and awareness-raising activities targeted at food insecure, climate vulnerable communities and national actors, and (ii)

Resilience building at the household and community level by diversifying livelihoods and establishing value chains for market access. Among the project's activities, targeted farmers will receive a cash transfer from the World Food Programme conditional on the completion (building or rehabilitation) of the community level adaptation assets identified during community consultations. Conditional cash transfers aim to support communities during the lean season in the short term and to increase resilience to extreme events in the long term. In this case, the proposal mentions an independent mid-term and final evaluation will be carried out. However, it does not include any discussion of how the causal impacts of the project will be estimated. Furthermore, the lack of clear target impact indicators in the proposal makes the project's overall impact goal vague and ambiguous. Boosting resilience can mean many different things, and at present, it is not explicitly clear what the project aims to achieve within the framework of increasing rural resilience.

Finally, reviewed FPs have vaguely discussed how well the programme will address the end beneficiaries and target the most vulnerable communities. A high quality M&E system can track if programmes reach their intended targets. Further, the programme logic model should indicate how it identifies and motivates target beneficiaries to participate.[10] Targeting is not as simple as calling for a certain percentage of participants to be women or poor, as many FPs do. Some FPs (20%) eligibility and targeting criteria are either not discussed or discussed but do not appear feasible given the programme design.

The funded project "Participation in Energy Access Relief Facility (EARF)" is of equal interest, implemented in 10 Sub-Saharan African countries. The COVID EARF is a sector-wide response to the economic downturn following the COVID-19 pandemic. The EARF's goal is to provide liquidity to energy access companies attempting to endure and recover from the health and economic crises that afflicted all countries. This financing aims at providing companies with the liquidity necessary to continue their operations, the flexibility needed to continue operations, and the ability to maintain their supply chain and inventory so they can grow post-crisis. Eligible companies must apply through the EARF website and meet the fund's eligibility criteria. However, there is no explicit discussion on the inclusion or exclusion criteria for identifying and selecting eligible companies and how the EARF will evaluate them to ensure compliance with GCF's standards and requirements. The proposal should have included selection criteria for beneficiary companies in the project's targeted countries.

---

[10] Nathan Fiala, Jyotsna Puri and Peter Mwandri, "Becoming bigger, better, smarter: A summary of the evaluability of Green Climate Fund proposals", Working Paper No. 1 (Songdo, South Korea, Independent Evaluation Unit, Green Climate Fund, 2019). Available at *https://ieu.greenclimate.fund/sites/default/files/document/working-paper-becoming-bigger-better-smarter-summary-evaluability-gcf-proposals.pdf*.

V. STOPLIGHT ASSESSMENT BY GROUP (INITIAL RESOURCE MOBILIZATION PERIOD VERSUS GCF-1 PERIOD)

This section presents a comparative analysis across the GCF-1 versus IRM period's funding proposals to assess how the quality at entry improves or changes over time. We intend to find any existing trends or patterns for funding proposals submitted before and after January 2020. Group 1, or initial resource mobilization (IRM) refers to proposals approved in the 2015–2019 period, while Group 2 or GCF-1 refers to proposals approved from January 2020 onwards.

A. STOPLIGHT ASSESSMENT

Starting with category A, we find that IRM/group 1 proposals are a **mix of low and medium risk**. Some 39 per cent of proposals are identified as low risk regarding ToC quality. Meanwhile, almost 49 per cent of proposals that include a discussion on the unintended consequences of their proposed intervention(s) are marked as low risk. However, 41 per cent of proposals in the IRM are considered medium risk in identifying causal pathways. On a related note, the robustness of the causal linkages is a greater source of worry as 32 per cent of proposals are high risk and 36 per cent medium risk.

By comparison, GCF-1/group 2 proposals are similarly a **mix of low and medium risk**, skewed towards the latter. Some 51 per cent of GCF-1 proposals are marked as medium risk on the quality of the ToC and causal pathways. Some 42 per cent and 32 per cent of proposals are estimated as medium risk concerning the robustness of the causal linkages and the quality of their supporting evidence, respectively. The areas of concern are unintended consequences, wherein 45 per cent of GCF-1 proposals are viewed as high risk while 21 per cent are high risk in terms of the causal pathways' robustness.

Table 6.   Stoplight assessment of the theory of change

	IRM			GCF-1		
A. Theory of change	% low risk	% medium risk	% high risk	% low risk	% medium risk	% high risk
What is the quality of the (implicit or explicit) theories of change and programme logic?	39%	38%	23%	34%	51%	15%
Are unintended consequences referred to and identified robustly in the programme ToC and/or in the surrounding literature reviews?	49%	33%	18%	8%	47%	45%
Are causal pathways clearly identified and discussed?	34%	41%	24%	40%	49%	11%
How robust are the causal linkages (implicit or explicit) and are they well informed by high quality evidence?	32%	36%	32%	38%	42%	21%
Is good quality evidence cited to discuss the efficacy of causal linkages?	49%	33%	18%	51%	32%	17%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

In category B, for IRM proposals, there are notable trends for all the indicators. Concerning the proposal design allowing for credible reporting of causal change, close to 50 per cent of proposals are low risk and 33 per cent medium risk. Similarly, in including activities pertaining to economic analysis and M&E, 48 per cent of proposals are graded medium risk, 24 per cent high risk and 28 per cent low risk. However, the trends that stand out are far more concerning. Three-quarters of IRM proposals are high risk regarding the adequacy of their M&E plans, while 68 per cent are in the high risk category regarding including methods for measuring causal change. Alarmingly, 93 per cent of all IRM proposals are at high risk of having biased estimates.

GCF-1 proposals are much more evenly risk-distributed. Including causal change measurement plans and strategies to reduce biased estimates remain the most concerning indicators, with 43 per cent of proposals being high risk for both these indicators. In terms of including activities about economic analysis and M&E, we see a very similar trend pre and post 2020, with nearly 50 per cent of proposals graded medium risk, 23 per cent high risk and 26 per cent low risk. Regarding the proposal design allowing credible reporting of causal change and having adequate M&E plans, 43 per cent are medium risk submissions, but 32 per cent and 23 per cent are respectively high risk.

Table 7. *Stoplight assessment of the potential for measurement of causal change and evaluability*

B. Potential for measurement of causal change and evaluability	IRM			GCF-1		
	% low risk	% medium risk	% high risk	% low risk	% medium risk	% high risk
Does the proposal design allow for credible reporting of causal change?	49%	33%	18%	25%	43%	32%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high quality evaluations and IEs?	11%	14%	74%	34%	43%	23%
What activities are included in the proposal that focus on 'economic analyses' and 'overall M&E' incorporated, and are these sufficient for high quality credible evaluations?	28%	48%	24%	26%	51%	23%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	9%	23%	68%	17%	34%	49%
Are there potential areas of bias that are likely to creep in?	3%	3%	93%	13%	43%	43%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

For category C, we see that IRM proposals are significantly low risk. Some 54 per cent of proposals are determined to be low risk in articulating eligibility and targeting criteria, while 83 per cent are low risk in implementation fidelity. Concerning measuring impact potential and performance against GCF investment criteria, approximately 80 per cent of IRM proposals are roughly split between low and medium risk. Notably, 46 per cent of group 1 proposals are medium risk in measuring performance against investment criteria.

We observe broadly similar trends for the proposals approved post-January 2020. Some 38 per cent of proposals are determined to be low risk in articulating the eligibility and targeting criteria, and 49 per cent medium risk. As concerns implementation fidelity, 66 per cent of all GCF-1 proposals are estimated as medium risk. For this group of proposals, 26 per cent are considered high risk in identifying and measuring impact potential, whereas 45 and 47 per cent are considered low and medium risk respectively for their paradigm shift potential. Lastly, 51 per cent of all GCF-1 proposals are deemed low risk in measuring performance against GCF investment criteria.

Table 8. *Stoplight assessment of implementation fidelity and performance against investment criteria*

C. Implementation fidelity and performance against investment criteria	IRM			GCF-1		
	% low risk	% medium risk	% high risk	% low risk	% medium risk	% high risk
Are eligibility and targeting criteria well articulated in submitted documents?	54%	27%	19%	38%	49%	13%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	83%	12%	4%	28%	66%	6%
To what extent is impact potential identifiable and measurable in the proposal?	43%	38%	19%	34%	40%	26%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	40%	31%	29%	45%	47%	8%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	33%	46%	21%	51%	43%	6%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Finally, in category D, we see that the proposals approved during the IRM period are skewed towards medium and high risk. Approximately 50 per cent of all proposals in this group are determined to be medium risk in terms of the adequacy of their M&E reporting plans, while 63 per cent and 56 per cent of IRM proposals are medium risk in terms of feasibility of measuring progress against investment criteria and the provision of additional impact indicators beyond the core GCF indicators, respectively. However, in terms of the adequacy and/or requirement of baseline data and the quality of data for rigorous IE, 49 per cent and 60 per cent of proposals are respectively estimated as high risk.

We see a much lower proportion of high risk proposals in the post-January 2020 group of proposals. Some 51 per cent of proposals are marked as medium risk regarding the adequacy of their M&E reporting plans, while 36 per cent are low risk. In terms of the feasibility of measuring progress against investment criteria, 45 per cent are low risk proposals, while 38 per cent are medium risk. This is reversed for the indicator on providing additional impact indicators beyond the core GCF indicators, where 32 per cent are low risk proposals while 49 per cent are medium risk. There are no high risk proposals in terms of the adequacy of baseline data, and 64 per cent of all GCF-1 proposals are viewed as low risk concerning their baseline data preparedness. Regarding the quality of data for robust IE, 53 per cent of all GCF-1 proposals are medium risk, and 32 per cent are low risk.

Table 9.   Stoplight assessment of data collection and reporting credibility

D. Data collection and reporting credibility	IRM			GCF-1		
	% low risk	% medium risk	% high risk	% low risk	% medium risk	% high risk
Are current reporting requirements sufficient for regular M&E?	21%	49%	30%	36%	51%	13%
How likely is it that progress on investment criteria can be measured credibly, given M&E plans, budget, and indicators for investment criteria?	19%	63%	18%	45%	38%	17%
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?	14%	56%	30%	32%	49%	19%
Has baseline data been collected and/or is there a requirement for this?	20%	31%	49%	64%	36%	0%
What is the potential quality of data and are these suitable for IEs?	10%	30%	60%	32%	53%	15%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

B. Conclusions

In conclusion, to better understand the pattern and trends of the GCF's funding proposals by IRM versus GCF-1 periods across four stoplight lenses we developed an analytical rubric of five scales (worsened/improved), with one being strongly worsened and five being strongly improved (Table 10). The average score by stoplight categories and analytical description is summarised in Table 11.

Table 10.   Scoring rubric

Scale	1	2	3	4	5
Description	Strongly worsened	Slightly worsened	No change	Slightly improved	Strongly improved

Table 11. Average score by stoplight categories

QUALIFIERS	THEORY OF CHANGE	POTENTIAL FOR MEASUREMENT OF CAUSAL CHANGE AND EVALUABILITY	IMPLEMENTATION FIDELITY AND PERFORMANCE AGAINST INVESTMENT CRITERIA	DATA COLLECTION AND REPORTING CREDIBILITY
Q1	4	**2**	**1**	4
Q2	**1**	4	**1**	4
Q3	5	3	**2**	4
Q4	4	4	4	5
Q5	3	5	5	4
Average scores	3.4	3.6	2.6	4.2
Assessment	Slightly improved	Slightly improved	Slightly worsened	Substantially improved

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

FPs submitted during the GCF-1 period slightly improved in articulating the theory of change and the potential for measurement of causal change and evaluability compared to the IRM period. In particular, GCF-1 FPs have significantly improved in identifying and articulating the causal pathways clearly. However, they are slightly worse at robustly identifying and discussing unintended consequences in the programme ToC. Likewise, GCF-1 proposals better identified and discussed their causal pathways and the potential areas of bias likely to creep in during the impact measurement. On the other hand, these FPs have slightly declined in their design and the likelihood of credibly reporting causal change.

We observe a slight decline in implementation fidelity and performance against investment criteria. GCF-1 FPs were generally scored at a marginally higher risk than IRM proposals. Primarily, GCF-1 FPs did not satisfactorily articulate the eligibility and targeting criteria nor support implementation fidelity with adequate and reliable information. However, GCF-1 did display a notable development in the likelihood of measuring and verifying performance against GCF investment criteria.

We also noticed a substantial improvement in data collection and reporting credibility among all categories in GCF-1 FPs. There was an enormous improvement in discussing planning and possibilities for baseline data collection. Similarly, given the M&E plans, budget, and indicators for investment criteria, a high likelihood exists for measuring and reporting the progress against investment criteria.

The evaluability study assesses the quality of GCF-1 funding proposals against IRM proposals across four categories: (i) the theory of change, (ii) potential for measurement of causal change and evaluability, (iii) implementation fidelity and performance against investment criteria, and (iv) data collection and reporting credibility. Compared to FPs submitted during IRM, those submitted during GCF-1 improved to some degree in all categories except implementation fidelity and performance against investment criteria. Data collection and reporting criteria demonstrated the most noticeable progress. This was evidenced in the substantial improvement in discussing planning and including baseline data collection possibilities in the proposal's annexes. GCF-1 FPs also showed a slight improvement in articulating the ToC and the potential for measuring causal change and evaluability. However, GCF-1 FPs slightly declined regarding implementation fidelity, which was not supported with adequate and reliable information. GCF-1 FPs also slightly declined regarding performance against investment criteria, where eligibility and targeting were not well articulated.

C. SUGGESTIONS FOR THE NEXT STEPS

From this assessment, we propose two main actions be followed by both the GCF and project proponents in order to ensure comparability and consistency in measuring and reporting climate change investments. These actions are essential to consider as many of the project proponents provide inconsistent information in their funding proposals, making it difficult to estimate the impact of the GCF investment credibly.

Low hanging fruit: Higher quality evidence is needed to satisfactorily discuss the efficacy of causal linkages when developing a project's theory of change. A need also exists for activities that help focus on 'economic analyses' and 'overall M&E' for high quality project evaluations. These will make it possible to pinpoint with supporting evidence the appropriateness of the GCF's investment and robustly report the likelihood of its impact.

Insightful, discerning actions: More measures should be implemented to support identifying unintended consequences when developing a project's theory of change. These measures could include synthesis and literature reviews. Furthermore, actions are needed to improve the implementation fidelity and performance against investment criteria. In particular, these actions should better articulate the targeting criteria and provide reliable information from the proposal regarding implementation fidelity. Lastly, more emphasis should be placed on how effectively the underlying impact potential is identified and measurable in the proposal. Therefore, it is imperative to focus on measuring investment criteria credibly by using M&E plans, budgets and indicators.

Appendix 1.   SUMMARY TABLES

Table A - 1.   Proposal distribution by sector

SECTOR	NUMBER OF PROPOSALS	PER CENT (%)	VOLUME OF FINANCE (GCF), USD	PER CENT (%)
Private	39	20.5	3,388.8 M	34.2
Public	151	79.5	6,529.7 M	65.8
Total	190	100.0	9,918.5 M	100.0

Source: GCF Tableau server as of 31 December 2021.

Table A - 2.   Distribution of GCF region by country focus

GCF REGION	NUMBER OF MULTI-COUNTRY PROPOSALS	NUMBER OF SINGLE-COUNTRY PROPOSALS	TOTAL
Africa	26	51	77
Asia-Pacific	13	66	79
Eastern Europe	6	5	11
Latin America and the Caribbean	15	31	46
Total	37*	153	190*

Source: GCF Tableau server as of 31 December 2021.
Note: * Double counting of proposals takes place as multi-country proposals can cover several regions.

Table A - 3.   Distribution of GCF region by theme

GCF REGION	ADAPTATION	CROSS-CUTTING	MITIGATION	TOTAL
Africa	36	19	22	77
Asia-Pacific	36	19	24	79
Eastern Europe	1	4	6	11
Latin America and the Caribbean	12	14	20	46
Total	82*	48*	60*	190*

Source: GCF Tableau server as of 31 December 2021.
Note: * Double counting of proposals takes place as some proposals can cover several regions.

Table A - 4.   Distribution of GCF region by sector

GCF REGION	PUBLIC	PRIVATE	TOTAL
Africa	53	24	77
Asia-Pacific	64	15	79
Eastern Europe	7	4	11
Latin America and the Caribbean	33	13	46
Total	151*	39*	**190***

Source: GCF Tableau server as of 31 December 2021.
Note: * Double counting of proposals takes place as some proposals can cover several regions.

Table A - 5.   Distribution of LDC/SIDS by country focus

COUNTRY CATEGORY	VULNERABILITY	MULTI-COUNTRY PROPOSALS	SINGLE-COUNTRY PROPOSALS	TOTAL
LDCs	Yes	25	52	77
	No	34	101	135
SIDS	Yes	17	28	45
	No	33	125	158
Total		37*	153	190*

Source: GCF Tableau server as of 31 December 2021.
Note: * Double counting of proposals takes place as some proposals can cover several countries/categories.

Table A - 6.   Distribution of LDC/SIDS by sector

COUNTRY CATEGORY	VULNERABILITY	PRIVATE SECTOR	PUBLIC SECTOR	TOTAL
LDCs	Yes	21	56	77
	No	31	104	135
SIDS	Yes	10	35	45
	No	37	121	158
Total		39*	151*	190*

Source: GCF Tableau server as of 31 December 2021.
Note: * Double counting of proposals takes place as some proposals can cover several countries/categories.

Table A - 7. *Stoplight assessment of the theory of change (by theme)*

BY THEME	ADAPTATION			CROSS-CUTTING			MITIGATION		
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?	40%	44%	16%	21%	37%	42%	29%	46%	25%
Are unintended consequences referred to and identified robustly in the programme ToC and/or in the surrounding literature reviews?	32%	46%	23%	53%	26%	21%	21%	40%	40%
Are causal pathways clearly identified and discussed?	35%	37%	28%	45%	45%	11%	31%	52%	17%
How robust are the causal linkages (implicit or explicit) and are they well informed by high quality evidence?	35%	44%	21%	39%	32%	29%	29%	35%	35%
Is good quality evidence cited to discuss the efficacy of causal linkages?	53%	35%	12%	66%	18%	16%	33%	42%	25%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 8.   *Stoplight assessment of the potential for measurement of causal change and evaluability (by theme)*

BY THEME	ADAPTATION			CROSS-CUTTING			MITIGATION		
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%	32%	16%	39%	37%	24%	25%	44%	31%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high quality evaluations and IEs?	19%	28%	53%	32%	21%	47%	10%	25%	65%
What activities are included in the proposal that focus on 'economic analyses' and 'overall M&E' incorporated, and are these sufficient for high quality credible evaluations?	28%	51%	21%	32%	53%	16%	23%	44%	33%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	16%	26%	58%	8%	29%	63%	10%	27%	63%
Are there potential areas of bias that are likely to creep in?	9%	16%	75%	8%	21%	71%	4%	19%	77%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 9.   Stoplight assessment of implementation fidelity and performance against investment criteria (by theme)

BY THEME	ADAPTATION			CROSS-CUTTING			MITIGATION		
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?	63%	28%	9%	45%	45%	11%	33%	35%	31%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	70%	25%	5%	74%	26%	0%	46%	46%	8%
To what extent is impact potential identifiable and measurable in the proposal?	42%	32%	26%	39%	47%	13%	38%	40%	23%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	35%	35%	30%	55%	32%	13%	40%	44%	17%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	46%	33%	21%	39%	50%	11%	33%	54%	13%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

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

BY THEME	ADAPTATION			CROSS-CUTTING			MITIGATION		
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?	26%	51%	23%	37%	53%	11%	19%	46%	35%
How likely is it that progress on investment criteria can be measured credibly, given M&E plans, budget, and indicators for investment criteria?	33%	49%	18%	37%	53%	11%	17%	60%	23%
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?	26%	49%	25%	24%	55%	21%	13%	56%	31%
Has baseline data been collected and/or is there a requirement for this?	33%	30%	37%	45%	29%	26%	33%	40%	27%
What is the potential quality of data and are these suitable for IEs?	21%	33%	46%	26%	29%	45%	8%	52%	40%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 11.   Stoplight assessment of the theory of change (by LDCs/SIDS)

By LDCs/SIDS	LDCs			SIDS		
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?	42%	46%	12%	32%	41%	26%
Are unintended consequences referred to and identified robustly in the programme ToC and/or in the surrounding literature reviews?	35%	37%	28%	38%	26%	38%
Are causal pathways clearly identified and discussed?	44%	39%	18%	29%	50%	21%
How robust are the causal linkages (implicit or explicit) and are they well informed by high quality evidence?	35%	44%	21%	26%	38%	35%
Is good quality evidence cited to discuss the efficacy of causal linkages?	54%	28%	18%	53%	26%	21%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 12.   Stoplight assessment of the potential for measurement of causal change and evaluability (by LDCs/SIDS)

By LDCs/SIDS	LDCs			SIDS		
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?	33%	44%	23%	35%	35%	29%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high quality evaluations and IEs?	16%	26%	58%	9%	32%	59%
What activities are included in the proposal that focus on 'economic analyses' and 'overall M&E' incorporated and are these sufficient for high quality credible evaluations?	18%	54%	28%	21%	56%	24%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	9%	26%	65%	3%	44%	53%
Are there potential areas of bias that are likely to creep in?	5%	21%	74%	6%	15%	79%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 13.   Stoplight assessment of implementation fidelity and performance against investment criteria (by LDCs/SIDS)

By LDCs/SIDS	LDCs			SIDS		
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?	46%	42%	12%	47%	38%	15%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	58%	37%	5%	59%	32%	9%
To what extent is impact potential identifiable and measurable in the proposal?	33%	46%	21%	44%	26%	29%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	44%	39%	18%	35%	44%	21%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	40%	49%	11%	35%	44%	21%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 14.   Stoplight assessment of data collection and reporting credibility (by LDCs/SIDS)

By LDCs/SIDS	LDCs			SIDS		
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?	26%	51%	23%	21%	53%	26%
How likely is it that progress on investment criteria can be measured credibly, given M&E plans, budget, and indicators for investment criteria?	26%	54%	19%	29%	44%	26%
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?	18%	60%	23%	21%	44%	35%
Has baseline data been collected and/or is there a requirement for this?	40%	35%	25%	41%	29%	29%
What is the potential quality of data and are these suitable for IEs?	16%	46%	39%	18%	32%	50%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 15.   Stoplight assessment of the theory of change (by country focus)

BY COUNTRY FOCUS	SINGLE-COUNTRY PROJECTS			MULTI-COUNTRY PROJECTS		
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?	34%	45%	21%	48%	34%	17%
Are unintended consequences referred to and identified robustly in the programme ToC and/or in the surrounding literature reviews?	33%	45%	22%	34%	14%	52%
Are causal pathways clearly identified and discussed?	32%	47%	20%	52%	31%	17%
How robust are the causal linkages (implicit or explicit) and are they well informed by high quality evidence?	33%	38%	29%	38%	38%	24%
Is good quality evidence cited to discuss the efficacy of causal linkages?	50%	35%	15%	48%	24%	28%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 16.   Stoplight assessment of the potential for measurement of causal change and evaluability (by country focus)

BY COUNTRY FOCUS	SINGLE-COUNTRY PROJECTS			MULTI-COUNTRY PROJECTS		
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?	44%	37%	19%	24%	38%	38%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high quality evaluations and IEs?	18%	25%	58%	28%	28%	45%
What activities are included in the proposal that focus on 'economic analyses' and 'overall M&E' incorporated and are these sufficient for high quality credible evaluations?	27%	51%	22%	28%	41%	31%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	12%	25%	62%	10%	34%	55%
Are there potential areas of bias that are likely to creep in?	7%	18%	75%	7%	21%	72%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 17.   Stoplight assessment of implementation fidelity and performance against investment criteria (by country focus)

BY COUNTRY FOCUS	SINGLE-COUNTRY PROJECTS			MULTI-COUNTRY PROJECTS		
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?	52%	32%	17%	34%	48%	17%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	66%	31%	4%	52%	38%	10%
To what extent is impact potential identifiable and measurable in the proposal?	42%	39%	18%	31%	34%	34%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	43%	35%	22%	38%	45%	17%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	39%	45%	16%	41%	45%	14%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 18.   Stoplight assessment of data collection and reporting credibility (by country focus)

BY COUNTRY FOCUS	SINGLE-COUNTRY PROJECTS			MULTI-COUNTRY PROJECTS		
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?	21%	52%	27%	48%	41%	10%
How likely is it that progress on investment criteria can be measured credibly, given M&E plans, budget, and indicators for investment criteria?	26%	58%	16%	38%	38%	24%
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?	21%	54%	25%	21%	52%	28%
Has baseline data been collected and/or is there a requirement for this?	35%	29%	36%	41%	48%	10%
What is the potential quality of data and are these suitable for IEs?	18%	34%	47%	17%	55%	28%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 19.   Stoplight assessment of the theory of change (by sector)

BY SECTOR	PRIVATE			PUBLIC		
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?	38%	44%	19%	37%	42%	21%
Are unintended consequences referred to and identified robustly in the programme table of contents and/or in the surrounding literature reviews?	41%	19%	41%	32%	44%	24%
Are causal pathways clearly identified and discussed?	47%	34%	19%	33%	47%	20%
How robust are the causal linkages (implicit or explicit) and are they well informed by high quality evidence?	31%	34%	34%	35%	39%	26%
Is good quality evidence cited to discuss the efficacy of causal linkages?	41%	28%	31%	52%	34%	14%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 20.   Stoplight assessment of the potential for measurement of causal change and evaluability (by sector)

BY SECTOR	PRIVATE			PUBLIC		
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?	28%	44%	28%	43%	35%	22%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high quality evaluations and IEs?	19%	25%	56%	20%	25%	55%
What activities are included in the proposal that focus on 'economic analyses' and 'overall M&E' incorporated and are these sufficient for high quality credible evaluations?	28%	44%	28%	27%	50%	23%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	13%	34%	53%	12%	25%	63%
Are there potential areas of bias that are likely to creep in?	3%	13%	84%	8%	20%	72%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 21.   Stoplight assessment of implementation fidelity and performance against investment criteria (by sector)

BY SECTOR	PRIVATE			PUBLIC		
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?	31%	44%	25%	53%	32%	14%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	56%	38%	6%	65%	31%	5%
To what extent is impact potential identifiable and measurable in the proposal?	38%	44%	19%	41%	37%	23%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	47%	38%	16%	41%	37%	23%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	44%	47%	9%	39%	44%	17%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 22.   Stoplight assessment of data collection and reporting credibility (by sector)

BY SECTOR	PRIVATE			PUBLIC		
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?	41%	28%	31%	23%	56%	22%
How likely is it that progress on investment criteria can be measured credibly, given M&E plans, budget, and indicators for investment criteria?	25%	53%	22%	30%	54%	16%
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%	50%	31%	22%	54%	24%
Has baseline data been collected and/or is there a requirement for this?	38%	34%	28%	36%	32%	32%
What is the potential quality of data and are these suitable for IEs?	13%	44%	44%	20%	37%	43%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 23.   Stoplight assessment of the theory of change (by AE modality)

BY AE MODALITY	DAE			IAE		
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?	39%	32%	29%	37%	46%	18%
Are unintended consequences referred to and identified robustly in the programme ToC and/or in the surrounding literature reviews?	32%	45%	23%	34%	37%	29%
Are causal pathways clearly identified and discussed?	39%	26%	35%	36%	49%	15%
How robust are the causal linkages (implicit or explicit) and are they well informed by high quality evidence?	32%	26%	42%	35%	41%	24%
Is good quality evidence cited to discuss the efficacy of causal linkages?	45%	35%	19%	51%	32%	17%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 24.   Stoplight assessment of the potential for measurement of causal change and evaluability (by AE modality)

BY AE MODALITY	DAE			IAE		
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?	55%	32%	13%	36%	38%	26%
To what extent are included requirements for M&E adequate and able to cover costs of undertaking high quality evaluations and IEs?	19%	26%	55%	20%	25%	55%
What activities are included in the proposal that focus on 'economic analyses' and 'overall M&E' incorporated and are these sufficient for high quality credible evaluations?	26%	48%	26%	28%	49%	23%
Are methods for measuring attributable causal changes (outcomes or impact or other) discussed?	10%	32%	58%	13%	26%	62%
Are there potential areas of bias that are likely to creep in?	3%	10%	87%	8%	21%	71%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 25.   Stoplight assessment of implementation fidelity and performance against investment criteria (by AE modality)

BY AE MODALITY	DAE			IAE		
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?	48%	35%	16%	48%	35%	17%
Is there adequate and reliable information included in the proposal regarding implementation fidelity?	68%	26%	6%	62%	34%	4%
To what extent is impact potential identifiable and measurable in the proposal?	32%	52%	16%	42%	35%	23%
To what extent is paradigm shift potential identifiable and measurable in the proposal?	42%	35%	23%	42%	38%	21%
How well are other GCF investment criteria informed and are these measurable and verifiable with high credibility and quality?	52%	39%	10%	37%	46%	17%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

Table A - 26.   Stoplight assessment of data collection and reporting credibility (by AE modality)

BY AE MODALITY	DAE			IAE		
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%	45%	26%	26%	51%	23%
How likely is it that progress on investment criteria can be measured credibly, given M&E plans, budget, and indicators for investment criteria?	23%	65%	13%	30%	51%	19%
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?	13%	68%	19%	23%	49%	28%
Has baseline data been collected and/or is there a requirement for this?	39%	29%	32%	36%	34%	30%
What is the potential quality of data and are these suitable for IEs?	19%	39%	42%	18%	38%	44%

Source: Assessments of funding proposals as of 31 December 2021. Assessment and analysis by authors.

GREEN CLIMATE FUND | Independent Evaluation Unit

TRUSTED **EVIDENCE**.
INFORMED **POLICIES**.
HIGH **IMPACT**.