



Evaluation Commissioned by FCDO

Evaluation of the Data for Foundational Learning (D4FL) Programme

Final Formative Evaluation Report
12 February 2025

In partnership with



Authors:

Alta Fölscher, Team Leader

Lydia Marshall, Deputy Team Leader

Christine Fenning, Evaluator

Rachel Outhred, Principal Expert

Paulina Valenzuela, Technical Expert

Tal Shalson, Research Coordinator

Sara Bashir-Malik, Researcher

Anushay Mazhar, Researcher

Mokoro Ltd

The Old Music Hall
106-108 Cowley Road
Oxford OX4 1JE

+44 (0)1865 403179
mokoro@mokoro.co.uk

mokoro.co.uk

Table of Contents

Executive Summary	iv
1. Introduction	1
1.1 Purpose and Scope of the Evaluation	1
1.2 Organisation of the evaluation report	2
2. Evaluation methodology and limitations	3
2.1 Overview of the approach	3
2.2 Methodology	3
2.3 Methodological limitations	6
3. The Data for Foundational Learning Programme and its Context	7
3.1 Context.....	7
3.2 The Data for Foundational Learning Programme	10
4. Evaluation Findings.....	15
4.1 Relevance	15
4.2 Coherence.....	23
4.3 Efficiency	29
4.4 Effectiveness	33
4.5 Looking ahead to impact and sustainability	39
5. Conclusions and Recommendations	43
5.1 Conclusions	43
5.2 Recommendations	47
Annex 1 Terms of Reference	50
Annex 2 Formative Evaluation Methodology	70
Annex 3 Formative Evaluation Framework.....	83
Annex 4 Original & Updated Theory of Change, Partner Theories of Change..	93
Annex 5 Use and Influence Plan	100
Annex 6 Timeline of Events in the Data for Foundational Learning Context.	102
Annex 7 Context for D4FL: Learning assessment in the context of SDG4.....	103
Annex 8 Potential Assessment Criteria of Progress.....	112
Annex 9 Assessing D4FL against the Evaluation Assessment Criteria.....	113
Annex 10 Explicated Recommendations.....	122
Annex 11 Bibliography	130
Annex 12 Acronyms	133

List of Tables

Table 1	Formative Evaluation Criteria and Questions	4
Table 2	Evaluation matrix for the formative evaluation	5
Table 3	MPLs for primary education	9
Table 4	D4FL implementing partners and roles.....	12
Table 5	Implementing partner processes to select countries for D4FL support	21
Table 6	Evidence of the delivery timeliness and efficiency to date.....	32
Table 7	Alignment between D4FL guiding principles and <i>Generate</i> and <i>Use</i> activities	35
Table 8	Formative Evaluation Questions	71
Table 9	Evaluation matrix for the formative evaluation	72
Table 10	Formative respondent selection for data collection.....	75
Table 11	Written outputs from the formative and summative evaluations	78
Table 12	Primary education MPL and equivalencies with cross-national assessments	106
Table 13	Availability of data to report against indicator 4.1.1 across regions	107
Table 14	Recommendations.....	122

List of Figures

Figure 1	SDG 4.1.1 Instrumental events	7
Figure 2	Countries supported by Oct 2024 through ACER UK, PAL Network and World Bank	14
Figure 1	Original D4FL Programme Theory of Change.....	93
Figure 2	Updated D4FL Theory of Change.....	95
Figure 3	ACER UK ToC mapped onto D4FL ToC	96
Figure 4	PAL Network ToC mapped onto D4FL ToC	97
Figure 5	World Bank ALMA ToC Mapped onto D4FL ToC	98
Figure 6	World Bank Accelerator ToC mapped onto D4FL ToC	99
Figure 7	Timeline of events.....	102

List of Boxes

Box 1	Selecting the country cases.....	74
-------	----------------------------------	----

Acknowledgments and disclaimer

Acknowledgements

The evaluation team expresses its gratitude to FCDO and partners for their contributions to the evaluation through interviews and workshops. The evaluation team is also grateful to Rona Bronwin, Matthew Harvey, and Willie Taylor, for their expert guidance to this evaluation.

Professional Quality Support was provided by Stephen Lister, to whom we are also thankful.

Disclaimer

The opinions expressed in this report are those of the evaluation team, and do not necessarily reflect those of the FCDO. Responsibility for the opinions expressed in this report rests solely with the authors. Publication of this document does not imply endorsement by FCDO of the opinions expressed.

The designations employed and the presentation of material in maps do not imply the expression of any opinion whatsoever on the part of FCDO concerning the legal or constitutional status of any country, territory or sea area, or concerning the delimitation of frontiers.

Executive Summary

The D4FL programme

1. **The Data for Foundational Learning (D4FL) programme aims to contribute to more low-income (LICs) and lower middle-income countries (LMICs) reporting internationally, and more state capability to regularly use disaggregated learning data.** It consists of three strands of work: (i) track global learning outcomes to enable reporting for SDG4 numeracy and literacy reporting and G7 girls' education targets (*Track*), (ii) generate more data through learning assessment tools, support with technical assistance, and promote citizen-led learning assessments (*Generate*), and (iii) ensure data is used to improve the planning, design and implementation of foundational learning (*Use*).
2. **D4FL is centrally managed by the UK-based FCDO unit *Education, Gender and Equality Directorate*, specifically the *Girls' Education Department*.** It is funded by FCDO. D4FL includes activities at the global, regional and country level, implemented with partners including UNESCO, the World Bank, the ACER International United Kingdom (ACER UK) and the People's Action for Learning (PAL) Network. The programme started operations in the UK Fiscal Year 2023/24 with a budget of GBP 27 million and will run to 2027.

The evaluation, methodology and limitations

3. **Mokoro Limited and Oxford MeasurEd are undertaking a mixed-method, Theory of Change-based formative and summative evaluation of D4FL.** The overall objective is to assess the extent to which – and how – the design and implementation of D4FL supports the achievement of its intended outcomes: (1) more LICs and MICs reporting internationally, and (2) more state capability to regularly use disaggregated learning data.
4. **The evaluation will cover all three strands of D4FL (*Track, Generate, Use*).** An external evaluation of the Global Education Monitoring Report (GEM-R) was conducted in 2023; therefore the current evaluation is not expected to re-evaluate the *Track* component, but to focus on its coherence and interaction with the other D4FL components. This scope and the timing of data collection meant that the formative evaluation focused on the design of the programme as a whole, and on the work delivered through three implementing partners: ACER UK, the PAL Network and the World Bank.
5. **The formative evaluation is focused on learning.** Its overarching questions are whether (i) each aspect of the programme is being implemented well, and working together to lead to the intended programme outcomes; and (ii) the FCDO and implementing partners could make changes to increase the likelihood of programme success.
6. **Formative evaluation data collection was undertaken between 1 October and 11 December 2024, following the methodology set out in the evaluation Inception Report.** This included document reviews and primary data collection with FCDO, implementing partners and independent stakeholders at global and country levels. We selected country cases to cover implementing partners in scope and a range of contexts: Ghana (World Bank, ACER UK), Kenya (PAL Network), Nepal (World Bank, PAL Network), Senegal (World Bank, PAL Network) and Somalia (World Bank, ACER UK).
7. **The findings, conclusions and recommendations presented in this report are limited by the depth of data collection possible within the evaluation timeframe and resources.** Factors are the number of country cases and number of country-level interviews undertaken to date, a notable reliance on interview and workshops with internal programme stakeholders and partner reports, and

the non-inclusion of countries who did not take up D4FL support. **We countered these limitations by triangulating evidence** and including respondents not directly connected to D4FL. We note where there is less confidence about specific findings due to these limitations.

8. **This formative evaluation report presents findings on early implementation of D4FL** and makes recommendations to improve programme design and implementation.

Evaluation Findings

Relevance

9. **D4FL's ultimate impact of improving foundational learning outcomes is relevant globally and across LICs and MICs.** There are low levels of foundational learning despite huge increases in enrolment, and evidence that a lack of foundational learning leads to poorer outcomes later in education and life. There is a lack of regular, comparable and disaggregated learning data.

10. **The objective of more LMICs reporting internationally on foundational learning is relevant given current SDG commitments and their monitoring.** Some global actors are concerned that the global momentum behind international reporting is waning.

11. **The objective of increasing state capability to use data to design and monitor foundational learning interventions is widely relevant.** There is broad consensus that state capability to use learning data at a national level is needed for countries to improve learning outcomes, but the barriers to this are multiple, including the availability of data, and the capability and political will to use it.

12. **The programme's design is relevant to achieving its objectives.** Each strand of work is relevant on its own. However, in conclusion we found that the balance of activities between strands may also hamper the equal achievement of both the programme's main objectives.

13. **Most D4FL-supported activities are oriented to supporting internationally comparable reporting,** taking advantage of in-country opportunities. This is the case for the activities by ACER UK, the PAL Network and the World Bank's ALMA. Programme stakeholders hope that, if countries supported by D4FL are able to report internationally, this will serve as "proof of concept" that will encourage more countries to take similar approaches and close global learning data gaps further.

14. **The TA being provided by ACER UK and ALMA includes strengthening in-country learning data systems,** even where the primary goal is to support countries to report internationally. While the work of the PAL Network under D4FL is focused on generating data for global reporting, the Network and its members see this contributing to more in-country demand for data and action, as well as providing data that is complementary to government data where it exists.

15. **To date, planned activities have not been focused on generating learning data in emergency contexts.** Generating and using learning outcome data in EiEPC contexts has a distinct set of barriers that are unlikely to be overcome without specific planning. While there is incidental inclusion of emergency contexts in country-level support, there are no specific activities under the D4FL for new tools or frameworks related such contexts, and the approaches being taken are likely not to be suitable for all EiEPC contexts.

16. **While all partners involved in the *Generate* and *Use* strands intend their work to result in data that is usable in-country, there are limited plans to support that use of data.** There is a particular gap in activities to strengthen use of simple foundational learning assessment in the classroom, which is set out as an intention in the D4FL ToC.

17. Current processes for identifying and meeting countries' needs are effective to meet immediate demand and leveraging partner expertise, but may not be fully strategic against the overall D4FL objectives. To date, country-level activity has been demand-driven, focusing on countries where a) the government has requested support for international reporting (ACER UK), b) World Bank country education projects and programmes, based on dialogue with country authorities, have identified an opportunity or need, or c) the PAL Network has identified capacity to roll out a national citizen-led data collection exercise. All of these reactive processes improve the likelihood of short-term success, but risk leaving out countries with significant but unarticulated needs.

Coherence

18. Partners share a unified aim of improving learning outcomes and agree on the importance of quality foundational learning data to drive action towards this ultimate aim. However activities in the *Track* strand are potentially contradictory. Different views among partners – for example on the value of supporting non-government data generation and on the merits of focusing on foundational learning vs. “learning poverty” – risk creating an incoherent narrative. Minimal interaction between partners to date has reduced opportunities for alignment.

19. While partners generally aim to avoid duplication, instances of perceived or actual overlap highlight challenges in achieving coherence. There are many countries where more than one D4FL partner is active, working on data generation. While there are efforts in some cases to undertake complementary work, timelines and the technical detail of the support provided may not be an easy fit. Coherency also depends in some cases on the capacity of country authorities to coordinate the work.

20. The *Generate* strand demonstrates strong potential to meet the needs of both the *Track* and *Use* strands. There is considerable optimism that the *Generate* strand’s potential to produce data that bridges global and national objectives, offering both international comparability and insights to inform in-country education action, and partners are focused on in-country needs for data.

21. However, tensions between global and local priorities, variability in planning and contextual challenges pose significant risks for this to be achieved. While generation of internationally comparable data has utility for the quality of data for in-country use, there are notable limitations to this utility. Limited planning up front for the use of data risks exacerbating this in-built tension between designing tools for global reporting and addressing country-specific needs.

22. Partners are actively working to align the logistical features of data generation exercises to ensure compatibility, and in some cases are integrating these with other assessment initiatives.

23. Partners intend to build on and strengthen existing in-country structures. Efforts towards this end include mapping existing assessments and data to avoid duplication (ACER UK), designing assessments or assessment frameworks with the aim of generating evidence that can inform and shape national education policies and reforms (ACER UK, ALMA) and generating citizen-led data that is complementary to and can be a check on government data sources (the PAL Network).

24. With this said, partners vary in their approaches to supporting government functions. The extent to which partners support the right functions within national assessment systems, and are able to achieve an effective and sustainable balance between TA and capacity building, remains to be seen.

25. D4FL by design complements other FCDO central programmes, and is designed to support and be supported by bilateral education programmes. At country level, this will

depend on advisors and different implementing partners leveraging this complementarity. Stronger D4FL mechanisms to promote joint working at country level, will benefit the programme.

Efficiency

26. **The D4FL commitment to work through country structures has mixed implications for timely and efficient use of resources.** Country pressure can contribute to fast use, but other factors like changes in leadership of counterpart institutions will cause delays. There is evidence that partners leverage positive factors, and take steps to mitigate negative factors.

27. **FCDO's decision to source partners directly has helped shorten timelines.** The potential loss of cost efficiency because of direct sourcing, is balanced by the opportunities for and practices of the FCDO and partners to maximise outputs for the amounts contracted.

28. **Partners appear to have internal processes that will help facilitate efficient use of D4FL resources,** including moving resources between countries to optimise value; running competitive processes to make on-rewards, and prioritising efficient use of resource. Longer partner timelines to use D4FL resources on the ground to date – for example where the World Bank and PAL Network have implemented their own processes to select countries– have been justifiable in terms of relevance and quality.

29. **Early evidence is that partners manage processes well, deliver on time and are efficient.** There are factors the facilitate efficient use of resources in the models of each partner, e.g. leveraging many-fold higher amounts in World Bank education projects and programmes, and the PAL Network having already invested in capacity building of its members.

30. **ACER UK is the only partner whose delivery model means directly responding to initial requests for support, and it has been swift to agree support and implement activities on the ground in line with country priorities and timelines.**

31. **While the number of strands, pathways within strands, partners and countries poses a risk that the programme spreads itself too thinly, this does not appear to have materialised to date.** Addressing risks to coherence and progress will maximise efficiency.

32. **The FCDO's management of the programme is robust relative to the management resources dedicated, and partners are satisfied with the support provided by the FCDO to keep work moving at pace.** However, the FCDO team is relatively small and there are trade-offs between oversight of implementing partners, providing the 'soft' support for coherence and coordination between partners, and undertaking global advocacy on learning assessments and the use of learning data in support of the programme and its objectives. The team would need to grow for the latter two to receive more attention.

Effectiveness

33. **There is some misalignment between the activities supported by D4FL and the original programme ToC.** Some posted ToC outcomes and outputs are not well supported, and some activities are not well articulated in the ToC, even where they contribute to programme impact or outcomes (e.g. the Accelerator). In addition, not all the outcomes and outputs are well supported by already contracted activities. Programme activities and investment is weighted towards *Track*, which may leave the programme short on demonstrating tangible results for *Use*.

34. **There is mixed evidence on the extent to which D4FL funded activities adhere to the FCDO's guiding principles for *Generate* and *Use* activities.** There is more evidence on support for government ownership and building state capability, than on including teachers and school leaders and providing cost-effective sustainable solutions.

35. Partners are making expected progress in delivering activities and outputs, and early results indicate that there will be at least incremental achievement on the programme outcomes within the programme timeframe. It is very likely that some countries supported by D4FL will be able to report against SDG4.1.1 by 2027. However, if the objective is that the programme results in a more systematic increase in the number of countries reporting, i.e. has an effect on countries other than those directly assisted by the D4FL, outcomes are less certain.

36. There is also some early progress towards generating better learning data nationally, but less on the use of data. The Accelerator programme is providing support to interventions aimed at improving curriculum and instruction, but the link to learning assessment and data is largely theoretical in the first crop of countries.

37. Political and institutional factors are likely to influence whether supported countries will go on to undertake regular assessments, track learning, and use the data for improving instruction and curriculum. These include political will, whether assessments are entrenched in education policies, country capacity, and assessment relevance. In addition, changing criteria for UIS eligibility and tensions about the extent of standardisation needed for global comparability are barriers specifically to countries reporting globally.

Impact and sustainability

38. The formative evaluation provides good pointers to pathways and criteria that can be used to evaluate progress to outcomes and impact. The evaluation team has proposed elaborations and clarifications to the original programme ToC, to guide the summative evaluation.

39. Partners' plans include some conceptualisation and integration of avenues to sustainable results, and a good understanding of the barriers, but sustainability can be integrated better. Partners are identifying and working towards opportunities for sustainable impact in-country, such as through capacity strengthening. However, sustainability could be better built into the design of the programme, including more efforts to secure buy-in at all levels and institutionalise demand for learning data in state decisions, to strengthen capacities to use data, and to reduce fragmentation in global and national learning assessment landscapes.

40. At a programme level, stakeholders identified an important opportunity for wider impact by using countries supported by D4FL to serve as proof-of-concept and encourage other countries to work on data generation and tracking. However, opportunities for this will be limited within the programme timeframe.

41. The FCDO does not currently have a clear influence strategy linked to the D4FL: the programme and sector can benefit from such a strategy. The FCDO can use specific and concrete examples of success from countries supported by the D4FL to influence the wider sector, through enabling sharing of information and advocacy at country level, regional level and global level. It is seen by stakeholders as a relatively neutral actor, a position it can leverage to promote more collaboration in adopting internationally and temporally comparable assessment frameworks.

Conclusions

42. Is each aspect of the programme being implemented well and working together to lead to the intended programme outcomes?

43. Some pathways are well designed individually, and early indication is that they are implemented well, such as the UIS/GEM-R *Track* and the ACER UK/ALMA and PAL Network pathways.

44. **Other pathways have significant design gaps.** The pathway to policies, targets and interventions using learning assessment to strengthen curriculum, assessment and instruction has few activities supporting it, and not necessarily in a strategic way. There are also some gaps in the *Track* and *Generate* pathways, particularly on how systematically the programme is contributing to assessment in EiEPC contexts, to improve learning data for affected populations.

45. There is, however, key design and implementation shortfalls on how well the programme is working together towards programme objectives. A significant issue for the programme, however, is how well the different strands are balanced and are working together, and whether the use of resources is optimal, to deliver programme results. The evaluation concludes three main things: (i) Programme investment is skewed towards *Track*, while *Generate* and *Use* is more relevant for impact. (ii) The integration of *Track* with *Generate* and *Use* is not well articulated to serve the programme objectives (iii) There is little scope for the FCDO to influence countries selection and interventions and not enough coordination between partners to ensure programming choices are strategic for the D4FL.

Recommendations

46. **Can the FCDO and implementing partners could make changes to increase the likelihood of programme success?** We have six recommendations, three about programme design and three about programme implementation.

Design recommendations

Recommendation 1. Adjust the ToC to reflect current programme design and intended mechanisms of change

Rationale: An accurate ToC will enable the summative evaluation to accurately evaluate the programme against its intended outputs and outcomes. It can serve as a programming tool, helping partners understand how strands of work fit together, and enabling learning conversations.

Proposed action: Adjust the Theory of Change to reflect all the currently intended D4FL inputs, outputs, intermediate outcomes, outcomes and impact, and how they are expected to lead to one another. Annex 4 Figure 2 provides an adjusted D4FL ToC as a base for discussion.

Responsibility and time frame: The FCDO, for immediate implementation.

Recommendation 2. Articulate the relative importance of national and global programme objectives

Rationale: The programme ToC weights filling global learning data gaps and providing disaggregated data for use at country level equal weight. In practice, there are often conflicts and trade-offs, however. There is also ambiguity about whether filling global learning gaps means UIS eligibility or any globally comparable data. A clear prioritisation of goals will bring clarity on where additional activities are needed and support better evaluation.

Proposed Action: Consider and decide/articulate more clearly the relative priority of countries having any useful data on foundational learning vs. meeting global reporting requirements.

Responsibility and timeframe: The FCDO, for immediate implementation.

Recommendation 3. Remove classroom use of assessment as an intended activity strand and shift to a focus on teachers as a key stakeholder group for D4FL *Generate* and *Use* activities

Rationale: The current ToC includes an outcome on strengthened use of simple learning assessment in classroom and community interventions. However, there are currently no planned activities to bring about this output. This would be a substantial new strand of work and would likely require bringing in new partners to ensure the right capabilities would be included.

Proposed Action: Reorientate thinking away from use of assessment in classroom and community interventions, towards prioritising teachers as a key stakeholder group in national assessment systems – as users of the data and experts in how children learn and what they need to know and be able to do in terms of foundational learning. This will have implications for D4FL programming, and requirements for/guidance on partners’ activities on the ground. We recommend that FCDO and relevant implementing partners consider early 2025 how teachers can be made more central in project efforts and what will be feasible within the time and resources available for the remainder of the project.

Responsibility and timeframe: FCDO, immediate; Partners, 2025-2027

Implementation recommendations

Recommendation 4. Increase the focus on the inclusion of EiE settings as a core programme objective

Rationale: The current programmatic ToC forefronts emergency contexts in intermediate outcomes and outcomes. But this was not identified as a particular focus for implementing partners. One solution would be to reorientate current efforts to ensure a more central focus on EiE contexts, but other global institutions hold the remit on this.

Proposed Action: Implementing partners should increase intentional focus on EiE; FCDO and partners should work together to develop a typology for assessing EiE contexts, to identify what will be feasible, which in-country stakeholders to work with, and what additional expertise might be needed. FCDO to link implementing partners to resources available on assessment in emergencies and to work with global partners on improving the architecture and evidence on learning data generation in emergency contexts.

Responsibility and timeframe: FCDO, immediate; Partners, 2025-2027

Recommendation 5. Improve joint thinking and cooperation vertically between FCDO and partners and horizontally between partners at the global and country levels

Rationale: A clear strategy by the FCDO on how it will influence programme direction, and closer cooperation between partners (including all Track, Generate and Use partners) will contribute to the programme’s strands working together better.

Proposed Action: The FCDO should think strategically about how it can influence all partners to pull together to make progress. FCDO and partners should follow through on ideas around shared principles on coordination and cooperation. The FCDO should facilitate more joint thinking and cooperation between partner at the global level, on how to achieve the programme objectives jointly, and address shared challenges. Partners would benefit from investing in this cooperation.

Responsibility and timeframe: FCDO on strategy and principles, immediate; FCDO and partners on joint thinking and coordination, 2025-2027

Recommendation 6. Partners should actively strengthen how they work politically to shift political and institutional barriers to regular assessment and use of data, and the FCDO and partners should invest in joint learning on how best to do so.

Rationale: Ensuring assessments are sustainable requires addressing political and institutional barriers to countries regularly undertaking assessments and using the resulting data to design and implement reforms in the instruction core of foundational learning. This recommendation is aimed at leveraging existing D4FL partnerships and contracted activities, to strengthen the links between Track – where a lot of investment is -- and Generate and Use, the engine room of learning improvement.

Proposed Action: Partners should strengthen whether and how their interventions will address existing and potential institutional barriers to imbedding learning assessments and using data, by shifting how assessments and learning data are seen and used. The FCDO and partners should invest in learning on how partners have successfully worked politically in this way.

Responsibility and timeframe: FCDO and partners, 2025-2027

1. Introduction

1. This is the draft report of the formative evaluation of the Data for Foundational Learning (D4FL) Centrally Managed Programme of the Foreign, Commonwealth and Development Office (FCDO). Mokoro Limited and Oxford MeasurEd are commissioned to undertake the evaluation between 2024 and 2027. This report is the evaluation's first main output; it provides findings on the early implementation of D4FL, and makes recommendations to improve programme design and implementation. The formative evaluation will be followed by a summative evaluation, done in two parts, a mid-term evaluation in 2025/26 and a final summative evaluation in 2026/27. This report is based on selected interviews at global and country level with the D4FL team, partners, sector experts and government counterparts, as well as a desk review of available documentation.

1.1 Purpose and Scope of the Evaluation

2. D4FL is a programme funded by FCDO and centrally managed by the UK-based FCDO unit *Education, Gender and Equality Directorate* (EdGE), specifically the *Girls' Education Department* (GED). By October 2024 D4FL included activities (planned or already under way) at the global level, in one region (the Caribbean) and at country level in 24 low-income (LICs) and lower middle-income countries (LMICs).¹ It is implemented with partners, including UNESCO, the World Bank, the Australian Council for Education Research (ACER UK) and the People's Action for Learning (PAL) Network. The programme has three strands of work: (i) track global learning outcomes to enable reporting for SDG4 numeracy and literacy reporting and G7 girls' education targets (*Track*), (ii) generate more data through learning assessment tools, support with technical assistance (TA), and promote citizen-led learning assessments (*Generate*), and (iii) ensure data is used by policy makers and partners to improve the planning, design and implementation of FL (*Use*). D4FL started in the UK Fiscal Year 2023/24, with focus on the mobilisation of six implementing partners to date and will run to 2027. The programme budget is GBP 27 million.

3. The objective of our evaluation of D4FL as set out in the TOR (see Annex 1) is to assess the extent to which – and how – the design and implementation of D4FL supports the achievement of intended outcomes:

- More LICs and MICs report internationally using comparable and disaggregated data on learning, including in emergency contexts.
- More state capability to regularly use disaggregated learning data so more governments and teachers implement and monitor interventions to improve foundational learning for all in priority countries.

4. The evaluation will cover all three strands of D4FL (*Track*, *Generate* and *Use*), using all six OECD-DAC evaluation criteria (Relevance; Coherence; Effectiveness; Efficiency; Impact and Sustainability). However, an external evaluation of the Global Education Monitoring Report (GEM-R) was conducted in 2023, therefore this evaluation is not expected to re-evaluate the *Track* component in full across all criteria, but to focus on its coherence and interaction with the other D4FL components. The formative evaluation therefore concentrates on three of the D4FL partners, the World Bank, ACER UK and the PAL Network, who were already contracted under the *Generate* and *Use* strands at the start of data collection in October 2024.

¹ The implementation partners were in the process of scoping and agreeing activities in a further 7 countries at the time of data collection.

Formative evaluation

5. The formative evaluation aims to provide evidence on whether and how the D4FL design and implementation are resulting in, or are likely to result in, progress towards the programme objectives. For this reason the evaluation occurred early in programme implementation. The overarching questions for the formative evaluation are:

- Is each aspect of the programme being implemented well, and working together to lead to the intended programme outcomes?
- Could FCDO and implementing partners make changes to increase the likelihood of programme success?

6. A full set of evaluation questions (EQs), linked to the OECD-DAC evaluation criteria, is set out in Table 1 below. The formative evaluation is intended to provide an in-depth understanding of how the different strands of D4FL are expected to work and to work together, to inform the final EQs and methodology for the summative evaluation.

1.2 Organisation of the evaluation report

7. This evaluation report is organised into five chapters, linked to annexes. These are
- **Chapter 2 – presents the methodology in summary**, with detailed EQs and approach to answering them. This chapter is linked to Annex 2 (which provides a detailed discussion of the methodology), Annex 3 (which provides the detailed Evaluation Framework) and Annex 4 0 (the original D4FL Theory of Change (ToC)). Annex 5 presents the evaluation use and influence plan and progress against it.
 - **Chapter 3 – provides a summary of the context** for the D4FL programme and a description of the evaluand itself. This chapter is linked to Annex 6 (a timeline of key global events in learning assessment tracking) and Annex 7 (a detailed context for the D4FL programme).
 - **Chapter 4 – presents the evaluation findings**. This chapter systematically answers the EQs, by criterion. It is linked to Annex 8 (additional tables) and Annex 9, which provides a detailed table of the EQs, sub-questions and assessment criteria, and RAG-rates² the programme against the assessment criteria, identifying which criteria were not covered by sufficient evidence to make a judgement. Annex 4, which also presents the explicated ToC for the D4FL and the partners, is also relevant to this chapter.
 - **Chapter 5 – presents the Conclusions and Recommendations**. The recommendations are presented in summary. Annex 10 provides more detail. The explicated ToC for the D4FL in Annex 4 is relevant to the recommendations.

² RAG means rating as 'red', 'red/amber', 'amber', 'amber/green' or 'green' where 'red' equals worst achievement.

2. Evaluation methodology and limitations

8. This independent evaluation began on 11 July 2024 with the consultative preparation of its Inception Report (IR) (Mokoro Limited and Oxford MeasurEd, 2024). Data collection and analysis for this draft was undertaken between 1 October and 11 December 2024, following the evaluation methodology set out in the evaluation IR. Annex 2 sets out the methodology followed in full, including information on the team and ethical considerations. There is no known conflict of interest concerning any team members. A summary of the methodology is provided below. Stakeholder analysis in inception (see Annex 2 for a summary) identified the audience and informants for the evaluation, and guided the team in selecting informants for the formative evaluation. See Table 10 in this Annex for people consulted, and Annex 11 for the main documentary sources and references. The evaluation team worked independently and without interference to produce the findings and conclusions of the report. It worked collaboratively with the FCDO and implementing partners on the recommendations.

2.1 Overview of the approach

9. This evaluation is a mixed-method, ToC-based evaluation. We take a realist approach, examining the context-intervention-result relationships at and between country and global levels.

10. Following consultations with the FCDO and implementing partners in the inception period, the team concluded that the existing ToC stood in principle, but proposed to work with FCDO and partners in the formative evaluation to expand the ToC, so that the summative evaluation tests a more detailed version (see Annex 4).

11. As set out in paragraphs 5 and 6, this formative evaluation is focused on learning. We set out to provide evidence on:

- the relevance and coherence of the D4FL programme design
- whether and how programme set-up and implementation to date are resulting in, or are likely to result in, effective, and efficient progress towards programme objectives
- potential pathways and opportunities for sustained Impact and Influence

12. The formative evaluation included desk-based research, primary data collection and analysis at portfolio and country levels. The country-level work will focus on the *Generate* and *Use* strands of D4FL but will consider coherence and interaction between these and the *Track* strand as relevant.

2.2 Methodology

Evaluation questions

13. We treat the overarching questions for the formative evaluation (see paragraph 5 above) as concluding questions, to be answered through considering individual EQs together.

14. In Table 1 below, we outline the 15 EQs for the formative evaluation, mapped onto the evaluation criteria. The EQs were refined with the FCDO team during the inception period. Due to the early stage of programme delivery and the formative nature of this evaluation, some questions are about what has happened to date, while some are about what is *likely to happen* based on how the programme and support have been set up.

15. While relevance, coherence, efficiency and effectiveness are the most relevant evaluation criteria for the formative evaluation, EQs 13-15 look ahead to likely impact and sustainability, with the intention of informing the final evaluation framework for the summative evaluation.

Table 1 Formative Evaluation Criteria and Questions

Criteria	Questions
Relevance	1. How relevant are the programme objectives to the needs of countries and the global community?
	2. How relevant is the programme’s design to achieving its objectives, including its equity objectives?
	3. To what extent are current processes likely to ensure countries’ needs are identified and met?
Coherence	4. To what extent is D4FL a coherent programme across strands and partners?
	5. To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies?
	6. To what extent are D4FL and other FCDO projects complementing one another?
Efficiency	7. To what extent will the programme set-up likely facilitate timely and efficient use of resources?
	8. To what extent are partners responding in a timely way to requests for support?
	9. To what extent have D4FL resources been used efficiently to date?
Effectiveness	10. To what extent is the implementation of D4FL activities likely to lead to outcomes as set out in the programme ToC?
	11. What has the programme achieved to date?
	12. What are the likely barriers and facilitators to achieving results?
Looking ahead to likely Impact and Sustainability	13. What are the likely pathways and timelines for outcomes and desired impact?
	14. What are the avenues for D4FL to have sustainable impact?
	15. What does sustainability look like for D4FL?

Data collection

16. Table 2 below provides an overview of the formative methodology and sets out the data sources we have used to address the evaluation criteria and questions. In Annex 3 we provide a comprehensive evaluation framework for the formative evaluation, outlining sub-questions, assessment criteria and data sources for each EQ. In Annex 2 we provide further detail on the nature and purpose of each data collection activity.

17. The evaluation included document reviews and primary data collection at global- and country-level. We selected country cases to cover all three implementing partners in scope for the formative evaluation, and include different regions, countries where there is/is not an FCDO education advisor, in different stages of implementation and where agreeing plans has challenging to date. We agreed with FCDO that instead of undertaking six cases (two per implementing partner), as proposed in the IR, we would undertake five cases but covering more than the two cases for two of the partners (World Bank and PAL). This allowed us to research the work of different partners in the same country, providing opportunities to look at coherence and synergies between partners’ approaches. The countries selected and agreed with FCDO were: Ghana (World Bank, ACER International United Kingdom (ACER UK)), Kenya (PAL Network), Nepal (World Bank, PAL Network), Senegal (World Bank, PAL Network) and Somalia (World Bank, ACER UK). Annex 2 provides the rationale for country selection, and Figure 2 in Section 3.2 maps the country footprints of the implementing partners.

18. Annex 2 Table 10 provides detail on primary research carried out at global and country level for the formative evaluation. At global level, we undertook 12 interviews with FCDO, implementing partners and independent stakeholders, as well as five workshops with the FCDO and implementing partners. At country level, we undertook 14 interviews with implementing partner representatives, government and other stakeholders, and FCDO advisors.

Table 2 Evaluation matrix for the formative evaluation

	Global-level					Country-level			
	Document review	External assessment expert interviews	TOC workshops with partners	Partner interviews	FCDO D4FL lead interview	Document review	Government interviews	Partner interviews	FCDO adviser interviews
Relevance									
1. How relevant are the programme objectives to the needs of countries and global community?	X	X		X		X	X	X	X
2. How relevant is the programme's design to achieving its objectives, including its equity objectives?	X	X	X						
3. To what extent are current processes likely to ensure countries' needs are identified and met?	X		X	X	X	X	X	X	X
Coherence									
4. To what extent is D4FL a coherent programme across strands and partners?	X		X	X	X	X	X	X	X
5. To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies?						X	X	X	X
6. To what extent are D4FL and other FCDO projects complementing one another?	X		X	X	X	X	X	X	X
Efficiency									
7. To what extent will the programme set-up likely facilitate timely and efficient use of resources?	X			X	X			X	X
8. To what extent are partners responding in a timely way to requests for support?				X		X	X		X
9. To what extent have D4FL resources been used efficiently to date?	X			X	X	X		X	X
Effectiveness									
10. To what extent is the implementation of D4FL activities likely to lead to outcomes as set out in the ToC?			X	X		X	X	X	X
11. What has the programme achieved to date?	X			X		X	X	X	X
12. What are the likely barriers and facilitators to achieving results?	X	X	X	X	X	X	X	X	X
Looking ahead to Impact and Sustainability									
13. What are the pathways and likely timelines for outcomes and impact?	X	X	X				X	X	X
14. What are the avenues for FCDO to have sustainable impact?	X			X	X	X		X	X
15. What avenues for FCDO to influence the wider sector?	X	X		X	X			X	X

2.3 Methodological limitations

Formative evaluation limitations

19. **Respondent bias:** the FCDO programme team and implementing partners might have an interest in overemphasising how programme activities are relevant, coherent, efficient, effective and sustainable, and a disincentive to divulge information on issues experienced. We tried to mitigate this by emphasising that implementing partners' own learning is an objective for the evaluation (enabled by including partners in evaluation reflection workshops); and by collecting perspectives from key informants who are not expected to or have not yet benefited from programme resources or inputs, such as external assessment experts and Government stakeholders.

20. **Limited sample of country cases and limited interviews at country level:** Evaluation resources mean that country case studies are limited, limiting the extent to cases will allow both sufficient comparison and variation (see paragraph 17 and Box 1 on sampling countries). Furthermore, in two of the four country cases where implementing partners had government counterparts, we were not successful in securing discussions with government representatives in the data collection period (Nepal and Senegal). We have tried to address this through triangulating country findings with partner engagements at the global level and with other stakeholders at country level. We also rated individual findings during analysis on the strength of evidence, noting where there is less confidence about findings, and communicating this in this report (see Annex 9). We intend to mitigate this limitation further in the summative evaluation through conducting a survey of all D4FL country stakeholders.

21. **Inclusion of countries who do not take up D4FL support:** Due to the limited timeframe of the formative evaluation and the early stage of programme delivery, the formative evaluation did not include countries who were offered or interested in the support but did not take it up. We addressed this limitation by analysing partner reports and asking interview questions about the level of interest and barriers to countries moving forward. The summative evaluation will include a survey of countries not taking up the support.

Lessons from implementation for the formative evaluation

22. The IR intended the learning and reflection workshops to include reflections on lessons for the summative evaluation. Because of the limited time we secured for the workshops, the evaluation team opted not to discuss these lessons and issues at this point, but to arrange time with the FCDO and partners in January 2025, to discuss the summative evaluation approach, framework and methodology. This evaluation report also therefore does not include a summary of lessons for the summative report.

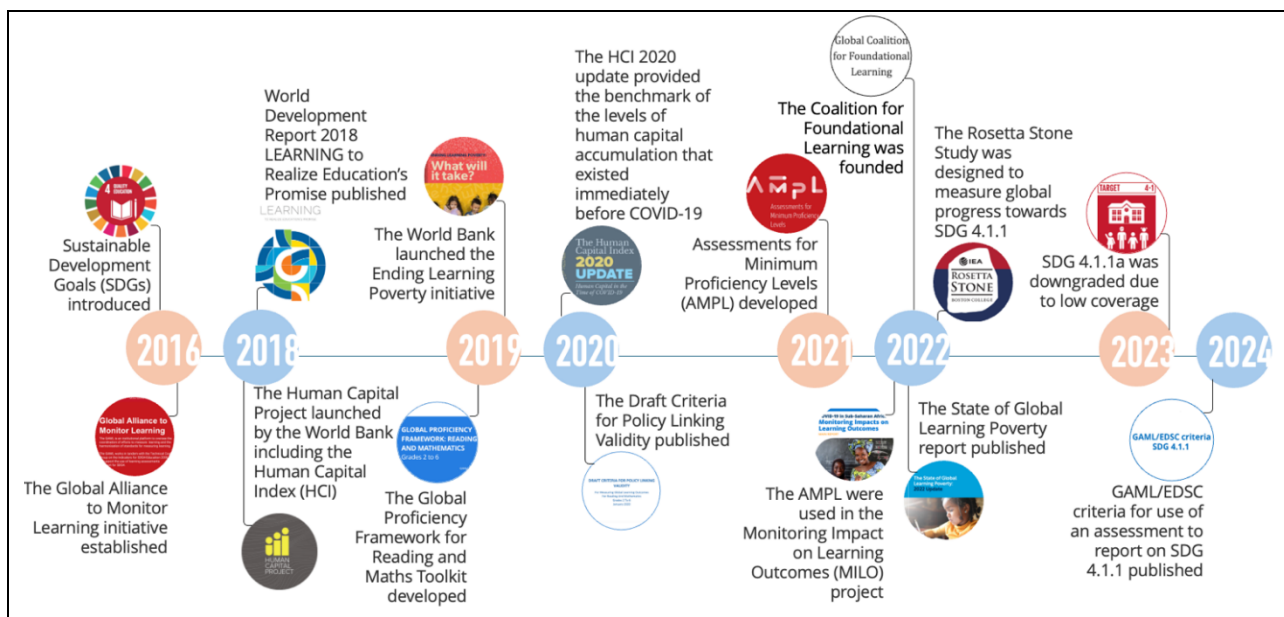
3. The Data for Foundational Learning Programme and its Context

3.1 Context

23. The IR included a comprehensive review of the background context for D4FL., which is reproduced as Annex 7 in this report. We provide a summary of that context below.

24. Figure 1 provides a timeline of instrumental events in the context of foundational learning data since the establishment of the Sustainable Development Goals (SDGs). A more detailed version is available in Annex 6 Figure 7).

Figure 1 SDG 4.1.1 Instrumental events



Source: D4FL Business Case and Authors of this report.

Political and technical caveats of assessment of learning outcomes for SDG4

25. SDG4 aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (UNESCO, 2015). It marks the first truly global agreement in education. For the first time, countries from diverse economic and cultural backgrounds agreed on a common set of educational objectives, demonstrating a collective commitment to improving education worldwide (UNESCO, 2017).

26. One of the most notable changes in SDG4 in comparison to previous education goals is its shift in focus from access to quality in education (Sayed & Moriarty, 2020). Unlike former global education initiatives that primarily aimed to increase access, measuring school availability, enrolment, completion, etc., SDG4 emphasizes the importance of ensuring that students are actually learning once they are in the classroom. By doing this, SDG4 acknowledges a critical reality: merely attending school does not guarantee that children are actually learning (Addey, 2024). One of the major outcomes of this shift has been the heightened importance of robust and reliable assessment systems to evaluate whether students are achieving the intended learning objectives.

27. While often presented as something purely technical, educational assessment is also deeply political. Methods, criteria, scoring mechanisms and dissemination of results will reflect worldviews, cultural norms, priorities, political will, and power dynamics. This political dimension was evident in the development of SDG4, which can be seen as a compromise between economics-driven and

human-rights-based perspectives (see paragraphs 2-5 of Annex 7), emphasising quality and inclusivity as an overarching goal while including specific targets addressing each view (UNESCO, 2015).

28. SDG4 has been broken down into 10 such targets, each with associated indicators, so that it can be monitored and assessed, ensuring accountability. However, operationalizing such a comprehensive goal and deciding on valid metrics for each of its aspects is not straightforward, and necessarily involves reducing complex concepts to measurable indicators. In the case of SDG4, the indicators chosen were influenced both by political aspects and by data availability. For example, target 4.1 represents a broad and ambitious aspiration: "By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes" (UNESCO, 2015). However, the indicators for this target refer only to proficiency levels in math and reading and completion rates in formal schooling. Because targets drive efforts towards measurable outcomes, often summarized in the principle "what gets measured gets done," it is likely that these indicators will guide focus and resources towards these important, but not all-encompassing aspects of learning, and away from the holistic nature of the overall goal.

29. Another challenge is defining proficiency levels (see Annex 7 paragraph 6). Because of vastly different educational realities, including systemic aspects, curricula, resource availability and teacher training, learning outcomes considered minimal in the global North may still be unattainable in many LICs and MICs. At the same time, governments might resist setting proficiency levels that would highlight widespread failure to meet targets, pushing to establish lower benchmarks. These disparities often complicate efforts to achieve a fair and accurate assessment of global learning.

30. Lastly, there are tensions around the choice of assessments to measure proficiency levels, even at the national level (see Annex 7 paragraphs 7-9). Ideally, this decision should respond to a clearly defined purpose and use methods that are aligned with it, as well as with resource availability.

31. A particular debate concerns the role of International Large-Scale Assessments (ILSAs), which are often perceived as the gold standard due to their focus on standardisation and statistical rigour, but with this view increasingly being questioned (see Annex 7 paragraphs 10-12). ILSAs are expensive, time-consuming, complex and require significant technical capacity and expertise. This has led researchers to call for and shift towards nimbler, cost-effective approaches that are aligned with the policy goals desired, giving due considerations for timing and resourcing constraints (see for example Wagner, Wolf, & Boruch, 2018).

32. Overall, we can see that while there is a formal consensus around SDG4, its implementation has revealed more nuanced challenges. Countries have agreed to the 2030 deadline, but there are still no clear accountability mechanisms to ensure adherence, and debates remain about measurement methods and consequences of failing to reach targets. Countries' willingness to align with certain perspectives and adapt their policies might often depend more on their varying dependencies on international organizations' influence than on their own SDG4 target commitment and their trust in assessment outcomes. Ensuring uptake, such that findings from SDG monitoring translate into political action and educational policy and practice, remains a significant challenge.

Availability of data for global monitoring of foundational learning

33. Of the 10 targets for SDG4, target 4.1 is particularly relevant for foundational learning. It aims to ensure that all girls and boys complete free, equitable, and quality primary and secondary education leading to relevant and effective learning outcomes. The most pertinent indicator for this target for D4FL is 4.1.1, which measures the proportion of children achieving at least a minimum proficiency level (MPL) in reading and mathematics at grade 2/3 and by the end of primary education (UNESCO,

2015). Math and reading were chosen as subjects because they are considered key within foundational learning, and metrics are more readily available than for other transferable skills.

34. Indicator 4.1.1 required defining what is meant by a MPL, which posed significant challenges. Curricula vary widely across countries, and cultural differences influence what is considered relevant. To establish global MPLs, several existing assessments were analysed to identify common performance level descriptors to guide a Global Alliance to Monitor Learning (GAML) consensus-building meeting in 2018 (UNESCO, 2018). This effort resulted in the global descriptors for MPLs outlined in Table 3 (see full table on MPLs mapped to different assessments in Annex 7 Table 12).

Table 3 MPLs for primary education

Domain	MPL	
	Grade 2/3	End of primary
Math	Students demonstrate skills in number sense and computation, reading simple data displays, shape recognition and spatial orientation.	Students demonstrate skills in number sense, computation, real world problems, basic measurement, 2D shape recognition, and reading and interpreting simple data displays.
Reading	Students read aloud and comprehend many single written words, particularly familiar ones, and extract explicit information from sentences. They make simple inferences when longer texts are read aloud to them	Students independently and fluently read simple, short narrative and expository texts. They locate explicitly stated information, interpret and give some explanations about the key ideas in these texts. They provide simple, personal opinions or judgements about the information, events and characters in a text.

Source: GAML, 2018

35. Despite these agreed definitions, some challenges persist. First, not all countries have collected learning outcome data, and those that have, have not necessarily done so at the target's grade levels. Even where data exists, aligning it with the global descriptors is complex. Much of the existing data was gathered before the MPL descriptors were agreed, or without using it to report against the MPLs being the main goal, leading to misalignments. For example, national assessments might prioritise monitoring progress towards curriculum-related learning outcomes, choosing to measure skills different from those in the MPLs. In other cases, assessments' highest possible performance levels are still lower than the MPLs, making them ineffectual for monitoring purposes.

36. According to the latest available data, approximately 70% of education systems have reported some data towards indicator 4.1.1. However, only about 53% of education systems have enough data to be able to report trends. It should also be noted that these percentages vary widely across regions: while 78% of education systems in Europe and Northern America can report trends, only 10% can in Oceania (Gustaffson, 2022) – for more detail see Table 13 in Annex 7.

37. At the same time the quality, reliability and alignment with MPLs of the data used for reporting differs substantially across education systems. Countries are instructed to prioritise data from international assessments like the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS), followed by data from regional assessment initiatives such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and Programme d'analyse des systèmes éducatifs de la Confemen (PASEC), and lastly national assessment data. However, some of the most common international assessments, such as Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA), as well as several regional assessments, were deemed by UIS to be unsuitable for reporting purposes for different reasons, including representativity, administration method and comparability (Montoya, 2024). Data coming from national assessments often suffers from similar issues.

Pending challenges

38. Annex 7 provides more detail on the following continuing challenges:

- **Exclusion:** The phrasing of indicator 4.1.1, which measures the percentage of children in school achieving MPLs in school, does not account for those out of school. Consequently, the indicator fails to capture the full picture of educational disparities and challenges, particularly among marginalized groups and in emergency contexts (see Annex 7 paragraphs 21-22).
- **Lack of uptake:** Challenges to uptake include the quality and accessibility of dissemination outputs and activities, a lack of concrete mechanisms for the coherent integration of data into decision-making and of decision-making into assessment design, and a (perceived or actual) lack of relevance of international measures to national contexts (see Annex 7, paragraphs 23-28). Despite these challenges, there is evidence of assessments having impact on policy, albeit most often through top-down changes. The areas most influenced by assessment results tend to be curricula and monitoring, evaluation and learning systems. Changes in more critical areas such as financing, pedagogy, or teacher training are much rarer, indicating a limited scope of influence for learning data on broader educational reform.
- **Misinterpretation of data:** Media coverage of learning outcomes can be hugely influential, swaying public opinion and prompting policy responses, but difficulty in interpreting results often leads to media outlets misrepresenting reports, leading to reactions that do not align with the actual findings (Annex 7 paragraph 26). This misalignment between the intended assessment messages and public perception can hinder effective use of data in shaping education policy.
- **Sustainability:** Many of the assessments deemed valid for reporting are both expensive and complex to implement. These costs and complexities are particularly burdensome for fragile states and regions experiencing educational emergencies. Efforts to build national and regional capacity in measuring foundational learning need to ensure that assessments are both affordable and practical, as well as strengthening system capacity, in order to lead to more reliable and actionable learning data (Annex 7 paragraphs 29-30.)

3.2 The Data for Foundational Learning Programme

39. The D4FL Programme is a FCDO-funded centrally managed programme (CMP), i.e. it is delivered by a UK-based FCDO unit. It is a successor to the Better Education Statistics and global Action to improve learning (BESTA) programmes.

Programme design and theory

40. The D4FL programme set out to deliver the following high-level objectives.

- **Impact:** Contribute to progress against SDG4.1.1³ and learning poverty reduction in LICs/MICs.
- **Outcome 1:** More developing partner countries report internationally on comparable and disaggregated data on learning, including in education in emergency contexts.
- **Outcome 2:** More state capability to regularly use disaggregated learning data so more governments and teachers implement and monitor interventions to improve foundational learning for all in priority countries.

³ SDG 4.1.1 Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level [MPL] in (i) reading and (ii) mathematics, by sex."

41. The original programme ToC is shown in Figure 1 in Annex 4. It shows how the D4FL programme was intended to achieve those high-level objectives through three strands of work: *Track*, *Generate* and *Use*.

42. The *Track* strand is focused on global tracking of learning outcomes, and specifically enabling LICs and MICs report against SDG4 foundational numeracy and literacy targets and G7 girls' education targets (*Track*). At design stage, there were two intended outputs of the *Track* strand:

- **Output 1:** Fund the generation, coordination and publication of comparable learning data using quality assessments.
- **Output 2:** Fund development and implementation of global methodologies to track foundational learning for all internationally.

43. Planned activities to achieve Output 1 included funding the PAL Network to generate internationally comparable learning data through the Network's citizen-led common assessment initiative, as well as continuing to UNESCO GEM to publish global reports, UNGEI/Cambridge Education to produce the G7 Global Objectives report, and UIS to track learning data. Output 2 was to include funding UIS to develop methodologies to measure and track learning outcomes, as well as the PAL Network's methodological development work.

44. The achievement of Outputs 1 and 2 was expected to lead to Outcome 1 through two intermediate outcomes: learning data gaps being filled with new data where none previously existed, and existing learning data being linked for global reporting.

45. The *Generate* strand is focused on supporting governments in LICs and MICs, with one intended output:

- **Output 3:** Fund TA to respond to national opportunities to adapt and implement rigorous regular primary learning assessments.

46. At the point of design, and through the course of this formative evaluation, the TA was to be provided by two implementing partners: ACER UK and the World Bank (through their ALMA programme).

47. The achievement of Output 3 was expected to lead to Outcome 2 by strengthening existing national assessment systems and education data. It was also expected to contribute to Outcome 1 in that it was expected that some countries receiving the TA would be supported to strengthen and/or link existing learning data so that it could be used to report internationally against SDG4 indicators.

48. The *Use* strand is focused on ensuring that learning data is used to improve the planning, design and implementation of foundational learning. It had two intended outputs:

- **Output 4:** Embed learning assessment in national policies and reforms on curriculum and teaching.
- **Output 5:** Strengthen use of simple foundational learning assessment in classroom and community interventions.

49. At the point of design, FCDO considered funding the PAL Network's work on assessment-informed action in the classroom through the D4FL programme. That work was in the end not funded by D4FL and, through the course of this formative evaluation, there were no specific activities aimed at achieving this output. TA provided by the World Bank Accelerator programme may include work on assessment in classroom and community interventions (e.g. formative classroom assessment), but it has not done so to date.

50. Planned activities to achieve Output 4 include the TA provided by ACER UK and the World Bank (through their ALMA programme), specifically where that TA included support to plan for the use of learning data in policy- and decision-making. It was also envisaged that the World Bank's Accelerator activities might include supporting governments with target setting and planning of Foundational Learning policy and practice informed by data and evidence.

51. The achievement of Outputs 4 and 5 was expected to lead to Outcome 2 by ensuring that policies, targets and interventions use assessment to strengthen curriculum assessment and instruction.

52. The original TOC did not set out explicitly how Outcomes 1 and 2 (i.e. more internal reporting on foundational learning and more state capability to use disaggregated learning data) would lead to the ultimate impact of improved learning. This causal chain was explored in the formative evaluation.

53. Based on findings from the formative evaluation, we recommended to adjust the Theory of Change (see Recommendation 1 in Section 5.2 of this report). Annex 4 Figure 2 provides an adjusted D4FL ToC as a base for discussion.

Programme implementation

54. Programme implementation commenced in 2023/24, comprising *Track* activities with UIS, GEM-R and UNGEI/Cambridge Education and mobilising the three further implementing partners. Activities with UIS, GEM-R, the World Bank, ACER UK and the PAL network occurred in 2024/25.

55. Implementation to date is summarised in Table 4 below, which also provides basic information about each partner, including their role in the programme, contracting date and share in programme funds. Figure 2 below shows the countries where partners are active. This progress is broadly as expected (see discussion under Finding 23).

Table 4 D4FL implementing partners and roles

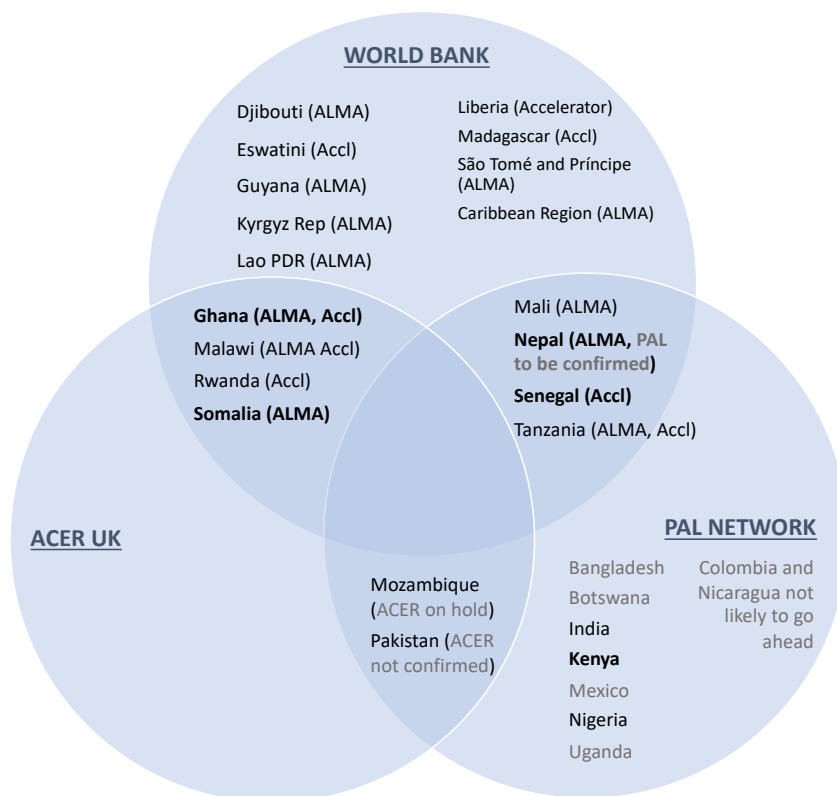
Partner information	Activities Implemented by end Q3 2024
Partially in Evaluation Scope (Coherence only)	
<p><i>Track</i> UIS, D4FL support to UIS to track and generate learning data. Contracted 3 March 2023, GBP 4.4 million</p>	<p>Global Coalition for Foundational Learning: UIS-led efforts on learning progress monitoring and expanding data availability through new tools and methodologies.</p> <p>AMPL-ab Project: Implemented in The Gambia, Kenya, Lesotho, and Zambia to measure lower and upper primary student proficiency. The report analysed SDG4.1.1a and SDG4.1.1b by gender across the four countries. Microsite with tools, guidelines, and publications developed, and AMPL visual identity created.</p> <p>Education Data and Statistics Conference: Held in Paris (7-9 February 2024, 420 participants, 131 countries and 80 organizations). Prior technical Cooperation Group (TCG) meeting and engagement day on lifelong learning.</p> <p>Releases: Education data released in March 2023 and September 2023.</p> <p>SDG 4 Scorecard: Second edition launched in collaboration with the GEM Report during the Education Data and Statistics Conference.</p>
<p><i>Track</i> GEM-R, D4FL support to GEM-R to provide evidence and analysis for policymaking. Contracted 3 March 2023, GBP 3.2 million</p>	<p>Research: GEM-R; SDG 4 Scorecard; Africa finance reports; gender-focused reports. Reports and background papers covered foundational learning, education leadership, climate change, and technology in education. Regional leadership studies initiated; 100 leadership profiles finalized.</p> <p>Communications and Outreach: Press releases and events; advocacy briefs and key reports launched at high-profile international conferences and regional events. Emphasis on promoting technology and gender equity in education.</p>

Partner information	Activities Implemented by end Q3 2024
<p><i>Track</i> UNGEI/Cambridge Education, D4FL funding to support the G7 Global Objectives report. Contract date not known, GBP 0.2 million</p>	<p>Report launched: Supported by UN Girls Education Initiative (UNGEI) and Cambridge Education, the 2023 Global Objectives Report was launched in February 2024 by the FCDO (UNGEI, 2023), part of a global initiative spearheaded by the UK and G7 leaders to improve access to and outcomes for girls' education. Dissemination: UNGEI has undertaken a webinar in February 2024 and an event at the Commission on the Status of Women 2024 on the report, and used its findings in the "Real Brave Change" campaign in June 2024.</p>
Fully in evaluation scope (All criteria)	
<p><i>Track and Generate</i> ACER UK, D4FL grant to support countries to develop quality, internationally comparable learning data. Contracted 20 Nov 2023, GBP 0.5 million</p>	<p>Scoping: Initial work to identify suitable countries January-March 2024 Ghana: From April 2024, support to National Council for Curriculum and Assessment (NaCCA) to align the National Standardized Test (NST), conducted in July 2024) with reporting requirements for SDG 4.1.1a and b. Work included: review of items and procedures with NaCCA; pre-work on UIS report; plans to support data analyses; work with NaCCA leadership to identify capacity building needs related to NST and UIS reporting; agreed co-designed NST2026 workshops. Rwanda: From April 2024, support the National Examination and School Inspection Authority (NESA) in preparing the Learning Achievement in Rwandan Schools (LARS) assessment for 2025 to report on SDG 4.1.1a and b. Work included: support for the development of a firm TOR to deliver the LARS in line with the UIS eligibility criteria; capacity building and training for local staff; support to understand the broader Rwanda assessment data landscape; and discussing actions to follow up on report recommendations. Mozambique: Support to Government, through the National Institute of Education Development (INDE), to enable the 2026 Avaliação Nacional de Aprendizagem (ANA) assessment to be used to report against SDG 4.1.1a. In addition, the request also included appropriate capacity building and training for local staff to ensure sustainability. It was agreed with FCDO on 29 July 2024 to pause work in Mozambique until after the November 2024 elections. Malawi: Work began in July 2024 to assist the Directorate of Quality Assurance and Standards (DQAS) in developing a unified vision for a revised assessment system to improve foundational learning. Recommendations from this work will guide potential future support under the grant. Somalia: From August 2024, support to the Federal education ministry Government Quality Assurance and Standards Department (QASD) for a National Framework for Assessing Learning Achievement. Work included: discussions with QASD; desk review; discussions key stakeholders; a virtual workshop in September to agree the intended scope and detail; and a draft framework. Framework intended for finalisation in January 2025.</p>
<p><i>Track, Generate and Use</i> PAL Network, D4FL grant for Citizen-led Common Learning Assessment to provide national level estimates of FL, comparable at a global level. Contracted: 01 July 2023, GBP 2.8 million</p>	<p>Development of tools and methodologies: PAL Network finalised a sampling framework tailored for the ICAN & ICARe assessments, aiming to capture reliable and comprehensive foundational learning data across multiple nations. Network developed a suite of monitoring and reporting tools. Country data collection: Agreements with network members to implement ICAN & ICARe assessments in 2025, and agreement with country statistics agencies on samples in: Tanzania, Kenya, Mali, Nigeria, Mozambique, Pakistan, India. Work with country statistical agencies has progressed for Bangladesh, Uganda, Mexico, Nepal and Senegal. Botswana is to be confirmed. Nicaragua and Colombia are likely not to proceed because of cost of data collection or not reaching agreement with partners</p>
<p><i>Track, Generate and Use</i> World Bank, D4FL funding for strategic activities for foundational learning through the</p>	<p>D4FL funding for additional TA to countries, alongside World Bank country education operations Accelerating Learning Measurement for Action (ALMA) supports countries to expand the availability and use of quality learning data by building on existing country capacities to implement learning assessments, close learning data gaps,</p>

Partner information	Activities Implemented by end Q3 2024
<p>Foundational Learning (FLC) Trust Fund. Contracted 1 March 2024, GBP 10 million</p>	<p>and strengthen assessment systems. ALMA works in countries that have not reported on SDG4.1.1. A call for proposals went out late June 2024 to 31 WB country teams in eligible countries as prioritised by World Bank regional Education Practice Managers. Proposals were received from 21 country education teams, and grants were made in September to Djibouti, Ghana, Guyana, Kyrgyz Republic, Lao PDR, Malawi, Mali, Nepal, Somalia, Sao Tome and Principe, and Tanzania, as well as to the Caribbean Region (Dominica, Grenada, St Lucia, and St Vincent and the Grenadines).</p> <p>Accelerator 2.0 has three components: country grants to support strengthening of instructional core (for example, through structured pedagogy programs) and related delivery systems; reinforcing regional technical expertise on foundational learning, including some support for high-level stakeholder convenings; and foundational learning knowledge creation and exchange.. A World Bank internal call for proposals went out end April 2024, based on a long list prepared by Africa Practice Managers. After internal review and scoring, 12 grants were made in September 2024 to WB country education teams to contract TA in Central African Republic, Cote d'Ivoire, Eswatini, Ghana, Liberia, Madagascar, Malawi, Republic of Congo, Rwanda, Senegal, Tanzania, and Sierra Leone.</p>
<p><i>Track, Generate and Use</i> NFER, D4FL funding for supporting a regional centre of expertise on learning assessment Contracted Nov 2024, GBP 0.2 million</p>	<p>FCDO and the National Foundation for Education Research (NFER) signed an accountable grant early November 2024 to support the Kenya Education Assessment Resource Centre to establish a regional centre of excellence covering assessment design, delivery, analysis and use. It intends creating a community/network of practice to share insights, best practice, existing materials/resources, and expertise.</p> <p><i>Note: this work is not included in the narrative above or current ToC as it was not live during the formative evaluation data collection.</i></p>

Source: FCDO (2022), (FCDO, 2024); Partner Reports; Interviews with the D4FL team and partners

Figure 2 Countries supported by Oct 2024 through ACER UK, PAL Network and World Bank



Accl = Accelerator Programme; Case studies in bold, Countries to be confirmed in grey

Source: Partner Implementation Reports, D4FL Management, Interviews

4. Evaluation Findings

56. The formative evaluation findings respond to the successive EQs from Table 1 above, and to the more detailed evaluation framework in Annex 3. In Annex 9 we RAG-rate the EQs and sub-questions against each assessment criterion.

4.1 Relevance

57. Relevance refers to the extent to which D4FL's objectives, design and activities to date respond to and meet the needs of LIC/LMIC country-level and global policy makers on and stakeholders in education outcome monitoring and standardised learning assessments.

EQ1 How relevant are the programme objectives to the needs of countries and the global community?

Finding 1. D4FL's ultimate impact of improving foundational learning outcomes is relevant globally and across LICs and MICs.

58. As detailed in section 3.1 of this report, most countries are not on track to achieve SDG4 by 2030, and it is projected that by that target year, 300 million students will not have the necessary skills in math and reading they need to succeed in life (United Nations Economic and Social Council. (2023). Moreover, progress appears to be slowing, a trend likely worsened by the Covid-19 pandemic and associated school closures (Engzell et al., 2021; UNESCO, 2023).

59. The SDGs highlighted the move from a focus on access to quality, and this continues to be the focus of the global sector. Global funders and institutions are reiterating the importance of foundational learning and the urgency of the learning "crisis", as demonstrated for example by the creation of the Foundational Learning Compact in 2020 and by the 2022 Commitment to Action for Foundational Learning.

60. The Commitment to Action for Foundational Learning has been endorsed by 35 countries. While there is clear evidence of the need for action on foundational learning, however, UNICEF's (2024) Learning Action Tracker suggests that most governments have policy priorities around foundational learning, but this is not always matched with system-strengthening action. Indeed, interviews with stakeholders for this evaluation echoed wider evidence that governments can lean towards focusing on enrolment rather than improvements in quality, because of the relative ease, though it is important to note that the two are not mutually exclusive.⁴

Finding 2. Increasing the availability of disaggregated learning data that can be used to design and monitor foundational learning interventions is widely relevant, as is increased state capability to use the data in this way.

61. For Government and other in-country actors to take action to improve foundational learning, they need to have evidence of learning levels and of the drivers of learning outcomes, and to be able to track progress over time. But in many LICs and MICs there is no existing, consistent source of

⁴ Evidence suggests there is a relationship, though not a "trade off" between the two policy objectives. See Pfeffer 2012, Crouch and Vinjevoold, 2006, and Spaul and Taylor (2013).

data on foundational learning at the national level. Examples from the country cases selected for this evaluation include:

- In Nepal, while national assessments exist, they are focused on later grades and reporting on earlier education is limited to enrolment rates.
- In Kenya, while various initiatives measure learning, government do not officially recognise many of them, and state reporting on foundational learning is limited to enrolment.
- In Somalia, there have been no national learning assessments to date.

62. Moreover, where data on foundational learning does exist (as in Ghana amongst the evaluation's country cases), it is not always of the required quality – for example lacking validity or reliability or not providing the particular information needed for decision making.

63. The availability of data is not sufficient on its own – other barriers to use of data include ownership, technical capacity, political will and resources to act on what the data shows. These challenges were reported at the global level and across the country cases for this evaluation. Indeed, one stakeholder interviewed for this evaluation reflected that supporting measurement is not particularly relevant if it isn't done alongside other processes to improve learning outcomes.

Finding 3. The outcome of filling global learning data gaps through more LICs and MICs reporting internationally is relevant to fulfilling SDG commitments and aligns with current monitoring systems.

64. As described by one stakeholder, the SDG agenda and focus on learning quality required a cultural and capacity shift, away from UIS statistics being based on administrative and survey data, to a need for a standardised way of collecting and reporting data on learning outcomes. While the SDGs are not legally binding, governments are expected to monitor progress against them. And yet, as detailed in section 3.1 of this report, most LICs and MICs are not currently reporting progress against 4.1.1, with the largest gaps for foundational learning (Gustaffson, 2022).

65. This formative evaluation provides a limited view on the perceived relevance of international reporting, given that it focused on internal programme stakeholders and countries that have sought out support from the D4FL programme. There are ongoing tensions in the global dialogue regarding the extent to which the sector should invest in global monitoring and reporting. The degree of measurement standardisation and comparability required to monitor learning against SDG4 is also contested, and has shifted multiple times during the SDG period (see Finding 26 paragraph 152). There are also ongoing epistemological questions regarding who should decide what children should know and be able to do as a result of schooling.

66. The D4FL has a clear theory on how international reporting can lead to the ultimate aim of improved foundational learning. The logic of the causal chain in the ToC (see Annex 4) is that highlighting low levels of learning and the progress that has or has not been made (and that can potentially be made), will lead to a focus on foundational learning and generate impetus, funding and/or demand for action. However, there was not a clear consensus across stakeholders on whether such value comes from comparability in any sense, or in formally recognised, globally comparable reporting against the SDGs.

67. While programme stakeholders emphasised the value of internationally comparable data for LICs and MICs, they recognised that there are varying levels of appetite for this among national governments. Reported barriers to government appetite for reporting internationally included:

- concerns about being shown to perform “poorly” compared to other countries
- political concerns around engaging with certain regional initiatives
- a view that international reporting requirements are imposed by international institutions and will not be relevant in-country

68. The hesitancy around being shown to perform “poorly” was not solely limited to internationally comparable data. Evaluation participants working on learning data generation perceived government hesitancy to be held “accountable” for low learning levels more generally.

69. Interview findings also suggested that the momentum behind international reporting might be waning, even where there was initial political will. Reasons for this included a sense that there has been a plethora of data generation initiatives that have not resulted in action or results, uncertainty around the criteria that will be used to evaluate whether data can be used for reporting against SDG 4.1.1, and a concern that those criteria will not be feasible in many low-income contexts.

70. To date, the D4FL programme has been demand-based and so, as mentioned above, this evaluation provides a limited view on the extent to which national governments will see the programme objectives as relevant to their contexts. However, there is evidence of countries approaching UIS seeking help with reporting internationally, which UIS were able to signpost to D4FL. There were other examples of the relevance of the programme objectives for national priorities too. For example, Kenya has just signed up to PISA, demonstrating the government’s continued interest in reporting globally,⁵ while the Ghana Education Sector Strategic plan identified reliable quality assessments as a priority for the country.

71. Additionally, programme stakeholders pointed to other actors interested in internationally comparable learning data. Examples given included development partners running education programmes in multiple countries and wanting to compare progress, and multiple partners wanting to roll out similar data generation initiatives as those backed by D4FL in specific countries.⁶

EQ2 How relevant is the programme’s design to achieving its objectives, including its equity objectives?

Finding 4. Most D4FL-supported activities, other than those funded by the World Bank Accelerator Programme, are oriented to supporting international comparable reporting.

72. Activities by UNESCO UIS, GEM-R, the PAL Network, ACER UK and the World Bank (through the ALMA initiative) are all oriented towards supporting more LICs and MICs to internationally report learning data. FCDO has deliberately decided to fund different ways of trying to achieve this goal:

- FCDO’s funding to the PAL Network has supported the development of a methodology to generate internationally comparable foundational learning data through citizen-led assessments, and will support data collection using this methodology in up to 15 countries.
- The TA being delivered by ACER UK, meanwhile, focuses in two countries so far (Ghana and Rwanda) on strengthening national assessments to align with UIS criteria so that governments can use their own assessments to report internationally. ACER UK’s current support for

⁵ Kenya previously participated in the UIS AMPL-ab project pilots.

⁶ One assessment criterion the evaluation does not cover in this discussion, is whether “Global learning assessment stakeholders have engaged with the D4FL programme and are interested in aligning activities and/or making use of outputs”. We have engaged with one external stakeholder, but have not collected other evidence as the programme is in too early stages of implementation.

developing country assessment frameworks in two additional countries, meanwhile, represents a pre-step to international reporting.

- The World Bank ALMA work is focused on providing technical and financial support to develop or strengthen national and regional learning assessments, to align with international standards and facilitate SDG reporting. ALMA is supporting eleven countries, of which three overlap with work by ACER UK and three with the PAL Network.

73. Programme stakeholders also hope that, if countries supported by D4FL are able to report internationally, this will serve as “proof of concept” that will encourage more countries to take similar approaches and close global learning data gaps further.

Finding 5. A particular issue in the equity orientation of the D4FL design is that, to date, planned activities have not been designed to leverage and address specific opportunities for and barriers to international reporting of learning data in emergency contexts.

74. Some of the countries involved in the programme include Education in Emergencies (EiE) context, and inclusion of emergency contexts may happen incidentally. Somalia, one of the country cases for this formative evaluation, is an Education in Emergencies and Protracted Crises (EiEPC) context, and this is one of the reasons that FCDO were keen for D4FL to operate in the country. To date the ACER UK support to Somalia has focused on first creating a national assessment system rather than supporting international reporting, but the World Bank’s ALMA support there will be aimed at supporting the implementation of the Assessment for Minimum Proficiency Levels (AMPL). The PAL Network is currently adapting tools for EiE contexts, with the aim to integrate this into future large-scale data collection exercises, but this will not be included in the 2025 round of International Common Assessment of Numeracy (ICAN)/ International Common Assessment of Reading (ICARe) data collection.

75. Generating and using learning outcome data in EiEPC contexts has a distinct set of barriers that are unlikely to be overcome without specific planning. For example:

- There is a variability of actors responsible for education service delivery, and often weak management of assessment processes within complex, multi-actor operating environments.
- Safety and security risks in many EiEPC contexts limit access for assessment teams.
- Identifying a sample frame is particularly difficult due to the high mobility of target populations and the lack of comprehensive population data, as well as the complexity of diverse formal and informal education provisions.
- Determining which curriculum to assess against or the language to assess in can be problematic. In some cases, multiple curricula are used across different age groups, or there is no officially recognised curriculum for refugee and internally displaced IDP children.

76. 2023 was marked by a record number of state-based conflicts, with 59 conflicts recorded, the highest number since data collection began in 1946 (Rustad, 2024). The need for education and data on learning outcomes in emergencies is therefore not only urgent, but growing rapidly (Outhred et al., 2024). School-based assessments, administered by governments through exams or sample-based assessments (national, regional or global) are often not implemented in EiE settings, making these children invisible in current data systems, including SDG 4.1.1 on learning (Outhred et al., 2024).

77. Implementing partners are aware that there are specific challenges in generating high quality data and in meeting international reporting requirements in emergency contexts, and have interest in doing so. But intentional focus will be needed to include such contexts in the D4FL work and the extent to which current partners have the required experience and expertise for assessment in EiEPC contexts is not clear.

Finding 6. The TA being provided by ACER UK and the World Bank’s ALMA include some strengthening in-country learning data systems, even where the ultimate goal often is to support countries to report internationally.

78. ACER UK’s primary aim in the D4FL programme is to support countries to generate quality data for international reporting, but with a secondary aim of strengthening in-country data and systems. For example, in Ghana, ACER UK’s support also builds the capacity of the NaCCA to undertake quality, reliable assessments. This includes capacity for item development, developing pre-assessment guidelines and reports, pilot testing and determining the psychometric properties of assessments. In Somalia meanwhile, as described above, the focus has shifted to first establishing a national assessment system that can later be strengthened to support international reporting.

79. Generating better information on the quality of education service delivery and resulting outcomes in country is more prominent in the ALMA programme. For example, in Nepal, the aim is to address specific issues with sampling and representativeness. In Ghana, the ALMA support includes building capacity of NaCCA to undertake quality, reliable assessments at the primary grade 2 level, and support for disseminating the results through a data portal. In Somalia, ALMA supports a small-scale data collection exercise using AMPL is planned as a back-up, so that if efforts towards nationally representative assessments “fail”, the government have some foundational learning data to inform education reforms (and report internationally). The Accelerator country grants can also fund assessment strengthening activities, e.g. in Senegal where the support is to strengthen EGRA in the local languages of two subnational regions.

80. Strengthening in-country learning data systems will also be the key focus of the new grant to NFER. Under the grant, NFER will provide TA to support the Kenya National Examination Centre to develop an assessment resource hub that will promote good practice in assessment to enable better reporting and use of data in Kenya and the region.

81. While the work of the PAL Network does not aim to strengthen state-run data generation, its member organisations see the ICAN/ICARE data as forming a part of the national assessment system. They conceptualise citizen-led assessments as not strengthening government processes, but strengthening the wider system by providing complementary data and fulfilling alternative purposes. For example, in Nepal, ICAN-ICARE data is expected to address specific gaps in current assessment coverage, including covering OOSC, incorporating private as well as government schools, being competency-based and focusing on foundational learning. Moreover, the intention of the work of the PAL network is to provide independent data that cannot be misrepresented by government and that can be used for advocacy and sensitisation (see Finding 12).

Finding 7. While all partners intend their work to result in data that is *usable* in-country, there are limited plans to *support* that use of data. A delayed focus on use may mean that data is ineffectively utilised.

82. The PAL Network member organisations that we interviewed aim to communicate findings to government and to citizens, but had no detailed plans for this at this early stage, with thinking about dissemination seen as “coming later” once the data had been collected. Similarly, in the ACER UK country cases the focus is on identifying the data that is needed for decision-making, but we have not heard about plans for supporting understanding and *Use* of data to date. In Ghana, ACER UK have identified a need to support government with how it *disseminates* results, but there are no concrete plans for this to date. The proposal for the new grant to NFER suggests that capacity building will cover “result analysis” and “reporting”, but it is not clear whether this will include a focus on analysing data for national policy decision-making as well as international SDG reporting. The World Bank ALMA programme objectives include strengthening the technical capacity of education ministries and assessment agencies in the use of learning data to promote improvements in education systems. In the country cases, references to “use” were limited to Ghana, where the ALMA grant will support capacity building for the NaCCA to report, and a portal to disseminate results.

83. In the Accelerator programme proposal mention is made that learning data, including that supported through ALMA, will support evidence-based identification of learning outcome shortfalls that can be improved through the Accelerator interventions (World Bank FLC, unpublished). In practice, only three of the 12 awards made by Accelerator in September 2024 explicitly mention in their summary descriptions using learning data to guide the activities funded. In interviews with World Bank stakeholders to date, data has been portrayed as playing a more general role in drawing attention to low learning levels and generating demand to improve what happens in the classroom, rather than providing specific information as to how interventions and education provision can be strengthened. In other words, while the Accelerator supports interventions to improve learning outcome, which in general would be based on learning data, in itself it does not directly support the utilisation of learning data.

84. Stakeholders warn that a delayed focus on use may mean that the data is ineffectively utilized. Indeed, learning from other assessment initiatives shows that planning for data use must be integrated from the start (see Annex 7).

85. As well as this lack of planned activity to support data use, programme stakeholders acknowledged that data on learning levels on its own is not enough for government and other decision-makers to know what needs to be done to improve those learning levels or solve the “problem” of learning poverty. For example, participants pointed towards the importance of background data (on socioeconomic and demographic characteristics, community and school factors) in illuminating the predictors or “drivers” of learning levels, as well as pointing to the schools and learners in most need of support. Additionally, participants expressed the need for further research to explain the “why” behind observed trends and disparities, including for examples case studies of high- and low-performing schools to identify lessons to inform policy and action.

86. There is a particular gap in activities to strengthen use of simple foundational learning assessment in the classroom, which is set out as an intention in the D4FL ToC (see Annex 4). There are currently no planned activities to support teachers’ use of data and/or assessment. Moreover, there is a lack of clarity around what the focus of any new activities to support this should be, and how the activities might be procured.

EQ3 To what extent are current processes likely to ensure countries' needs are identified and met?

Finding 8. The processes for deciding who should support whom, and how, under the D4FL initiative are effective in meeting immediate demand and leveraging partner expertise, but may not be fully strategic against the overall D4FL objectives.

87. There are good examples of the FCDO and D4FL implementing partners responding to government requests for support to meet UIS requirements (e.g. Ghana, Rwanda and Malawi), and aligning with national activities (see Section 4.2), as well as capitalising on a general demand for data in-country. Generally, because the FCDO's selection of implementing partners has been based on their credibility, reputation and experience, the programme is deploying partners with appropriate capacities to meet specific country needs.

88. However, the processes for deciding who should support whom, and how, are less likely to *comprehensively* meet LICs' and MICs' needs due to variability in selection processes and criteria between partners and the limitations of demand-driven approaches.

89. The D4FL TOC and business case provides a clear rationale for the selection of current partners and their roles in implementing the programme, as well as a framework for evaluating new partners and country opportunities. The TOC was formulated to allow scope for funding convincing opportunities that will strengthen assessment, the availability of learning data and its use, given the dearth of such data, the technical challenges of quality assessments, and large gaps in education practice on using learning data to strengthen teaching. It is not clear to the evaluation that the TOC and business case, as they stand, are sufficient to ensure that the net effect partner and country selections by the end of the programme, will achieve programme objectives optimally.

90. Firstly, processes to select countries to support are partner specific, and while this means that partners have the right expertise and relationships to enhance their relevance to specific countries for specific purposes, across countries and within countries there may be gaps.

Table 5 Implementing partner processes to select countries for D4FL support

ACER UK	Country selection has been responsive to the demands of specific countries, based on requests from governments (articulated through UIS) or the presence of an FCDO advisor in-country. This strengthens relevance to the country, but may not strategically assist FCDO to provide proof of concept. In this case the FCDO directly influences which countries are supported.
PAL Network	The ICAN/ICARE initiative will only include countries where network members have implemented previous citizen-based assessments. The final selection of countries will be based on readiness, including progress in obtaining permissions and securing resources, often leveraging existing networks. This ensures feasibility but might exclude countries with less-developed systems that require foundational support. This approach to country selection was approved by FCDO as part of the PAL Network's proposal.
World Bank	A thorough proposal process was used to assess country context, government appetite, and potential to accelerate improvement in learning outcomes through World Bank lending programmes. While this ensures alignment with broader World Bank initiatives, the link to lending cycles and funding constraints limit support to countries with urgent needs, but less strategic from a FLC perspective, needs. The FCDO has no direct influence over which countries are selected.

91. Secondly, FCDO's reliance on partner credibility and demand-driven support risks perpetuating an *ad hoc* approach that may not fully address systemic gaps or prioritize equity across regions. For

example, prioritizing technical assessment expertise in implementing UIS-compliant assessment for global tracking, has sidelined other critical needs, such as strengthening national systems, engaging data users, or addressing emergency contexts. It is not that these needs are ignored in the programme overall, but whether they will have been met adequately by the end of the programme.

92. While responding to demand increases the likelihood of success, it may constrain the programme's ability to achieve systemic, long-term impact. Demand-driven models may overlook countries with significant needs but low visibility or weak systems, particularly those in EiEPC contexts. It also raises questions regarding the scalability of any "proof of concept" of models to generate globally comparable data, but which relies on specific partners.

93. There are counterarguments for these concerns: The PAL Network's wide reach supports the possibility of building momentum across regions. The use of ACER UK reflects the belief that early successes in demand-driven contexts can generate interest and expand impact through awareness-raising activities. However, ACER UK stakeholders acknowledge that the demand-driven approach taken to date improves chances of success but limits broader systemic impact. A more proactive strategy might help address needs in countries with less visible demand but significant challenges.

94. Thirdly, while the TOC and business case provides a framework for evaluating new opportunities for support and deciding what and who to fund with the uncommitted portion of the D4FL budget, selection of new partners is at any time in response to new opportunities, which is not the same thoroughly appraising all options and selecting the best, for optimal achievement of outcomes. This poses a challenge to ensuring long-term coherence and alignment, as well as optimising relevance. This challenge could be addressed by being proactive about opportunities, and making sure that the availability of funds to support foundational learning measurement and use are well known beyond the programme partners.

Finding 9. Partners' processes for identifying and ensuring that TA meets countries' needs demonstrate promising early results, with evidence of tailored scoping, contextualization, and strong partnerships.

95. Although it is early to comprehensively assess the adequacy of all partners' processes, initial evidence suggests that partners are taking steps to align TA with countries' needs. For example,

- ACER UK has conducted detailed scoping exercises to understand country-specific needs and align their support. In Ghana, ACER UK worked closely with NaCCA to assess capacity-building needs, ensuring alignment with both global reporting requirements and national data usage goals. In Somalia, ACER UK have tailored their support to first support the QASD to develop a Learning Assessment Framework, in response to the ACER UK scoping and needs assessment.
- The PAL Network emphasizes contextualization by adapting assessments to local cultural, structural, and linguistic differences. For instance, in Nepal, the member organisation collaborated with government departments and universities to ensure alignment with national priorities. In Kenya, the member organisation will work with subnational partners and local authorities to manage data collection, leveraging established relationships and processes.

96. These examples highlight partners' efforts to ensure D4FL activities are relevant and responsive to local needs. However, despite positive efforts, challenges remain. These include variability in contextualisation and reliance on local stakeholders. For example, while the PAL Network emphasizes contextualization, it remains unclear how extensively tools will be tailored beyond the language of

assessment. This raises questions about the depth of alignment with unique local priorities and systems. The PAL Network's model also places significant responsibility on local members to engage in further partnerships. In contexts with weaker systems or limited capacity, this could result in inconsistent implementation.

97. Evidence of the actual relevance of TA and alignment to countries' needs is limited to date, given the early stage of programme implementation (see Sections 3.2 and Finding 14 of this report). There are however positive examples emerging:

- In Ghana, the planned interventions align with Ghana's need to report internationally, and stated priority to use assessment data to strengthen basic education. ACER UK's support to date is viewed by government to be relevant to their needs, and to have already contributed to better quality of assessments for national purposes.
- In Somalia, ACER UK is working closely with the QASD to develop the Learning Assessment Framework, aligned to the Somalia National Education Framework. The consultations are in Somali, to ensure accessibility. The QASD has confirmed their satisfaction with the support.
- The World Bank operations are implemented through government which should facilitate alignment with need. However, as the support has not yet started, the evaluation could not fully confirm the relevance of the TA. In the country cases, however, there is some evidence of relevance, given alignment in principle with articulated country needs (e.g. Ghana, Somalia) or stated policies (e.g. Senegal and Nepal).

98. The evaluation country cases found no evidence of non-alignment with country needs.

4.2 Coherence

99. Coherence refers to the extent to which D4FL's objectives, design and activities to date are internally coherent and complementary, and align with country- and global-level education structures, policies and priorities and assessment initiatives.

EQ4 To what extent is D4FL a coherent programme across strands and partners?

Finding 10. D4FL partners demonstrate a strong alignment in their overarching objectives of improving foundational learning and generating quality learning data. They also aim to avoid duplication but there are instances of overlap, suggesting scope for better coordination, both globally and in-country.

100. Annex 4 shows how the work of each partner maps onto the programmatic theory of change, aiming to contribute to this ultimate impact. The coherent focus on generating quality learning data is demonstrated by a collective and consistent emphasis on producing quality data that is "fit for purpose" and can inform both global standards and national education policies.⁷

101. An exception to this alignment is the Accelerator initiative, which shares the ultimate aim of D4FL of improving learning but focuses on strengthening teaching practices, government systems, and curriculum reforms rather than using data as the primary pathway. The intention is that the

⁷ On the evaluation framework criterion regarding countries share this understanding, this is not assessed in the section as we do not know enough yet about whether countries share the understanding.

work through the Accelerator initiative makes use of learning data, but as discussed above (see paragraph 20) this link has not been given high priority to date.

102. Despite shared objectives, several internal challenges hinder full alignment among partners:

- **Different indicators:** Some partners emphasize SDG 4.1.1a foundational learning metrics, while others prioritize the World Bank's learning poverty metric. Critics argue that this difference risks diluting efforts toward achieving SDG 4.1.1a, as what gets measured often dictates funding and focus.
- **Tensions around global standards:** Coherence within the *Track* strand is threatened by debates over global certification requirements. For instance, D4FL funds both the UIS to set standards and partners who develop methodologies and generate data that UIS may not go on to certify. While some view diverse data tools as complementary, others see this as a lack of a unified framework, complicating the effort to generate consistent, actionable insights, and detracting from having as many countries as possible report.
- **Lack of consensus about the importance of working within government systems:** The PAL Network operating outside government systems may limit its contribution to strengthening national capacity, but allows it to address gaps in accountability as well as providing an option for data generation where there is not state capacity to do so. However, there is a lack of agreement among project stakeholders about the value of working outside of government, with some stakeholders believing that working outside of government structures in this way undermines long-term system strengthening. This lack of agreement could lead to incoherence or contradiction in external conversations about the D4FL programme.
- **Limited collaboration:** Links between partners remain sparse at both global and in-country levels. Examples of collaboration, such as coordination between ACER UK and the World Bank in Ghana, are promising but rare. In many cases, partners are unaware of overlapping activities, leading to risks of duplication or missed opportunities for complementary work, as seen in Nepal and Senegal.

103. Contextual factors can also make coordination between partners difficult. For example, in EiEPC settings, there are often multiple parties operating in silos. In Somalia, stakeholders identified that weak donor coordination and overlapping initiatives have created risks of duplication. In that country case, the World Bank ALMA programme and ACER UK's work on national assessments must articulate how their efforts align, and align with the Ministry of Education's vision, to ensure coherence.

104. While partners generally aim to avoid duplication, instances of perceived or actual overlap highlight challenges in achieving coherence. The World Bank's ALMA programme and ACER UK's work are both aimed at supporting countries to generate learning data. Both initiatives are undertaking activities in three of the countries supported by D4FL (Ghana, Malawi and Somalia), working with the same counterparts. In Ghana, for example, both initiatives aim to support the NaCCA on analysis and reporting to UIS. While at the detailed level the work appears complementary (working with assessments at different grade levels – primary level 2 versus levels 4 and 6) there is the potential that the advice the two partners give to the NaCCA may be contradictory. In Somalia, while ACER UK is currently working on a Learning Assessment Framework, this may extend into supporting assessments, which ALMA is also aiming to do. In both countries the partners are speaking or aiming to speak, but complementarity of support also depends on country authorities' ability to integrate the work of different partners.

105. Another concern raised about duplication was the risk that the PAL Network initiative might duplicate efforts in countries where governments are working towards using national assessments for global reporting. Even in Nepal, where there is no immediate plan to report against SDG 4.1.1 using national assessment data, there were initial concerns from government that ICAN/ICARe would be duplicative of other data. These concerns were overcome through dialogue, however. As discussed in Finding 12 paragraph 116, the PAL Network's citizen-generated data can be complementary to state-generated data. In Kenya and Senegal, stakeholders identified opportunities for added value through different partners generating complementary data. However, this concern points to the importance of dialogue and collaboration between D4FL partners and between the PAL Network and other in-country stakeholders.

106. Addressing these issues will require systematic coordination, clearer alignment on priorities, and mechanisms to foster both global and in-country collaboration. There is significant potential for D4FL partners to work more cohesively in-country. In some contexts, partners are beginning to align their work, as seen in Kenya and Somalia, where they are exploring how their activities might complement one another. The Ghana example demonstrates the value of pragmatism, with one partner's work (ALMA) picking up where another's ends (ACER UK).

107. Partners agree on the need to systematize collaboration to prevent duplication, maximize complementarity, and present coherent narratives to governments. However, there is no formal requirement or mechanism to facilitate this coordination. Concerns were also raised about instructions to collaborate in-country adding burdens or appearing patronizing to in-country colleagues. FCDO advisors could play an important facilitative role in linking up different partners and strands of work in country, but this depends on individual interest and capacity.

Finding 11. The *Generate* strand demonstrates strong potential to meet the needs of both the *Track* and *Use* strands by producing data that is relevant for both global reporting and in-country application. However, tensions between global and local priorities, variability in planning, and contextual challenges pose significant risks in this regard.

108. There is considerable optimism that data generated through the *Generate* strand can serve both global and in-country needs. For instance:

- ACER UK aims to support countries in using their national assessments for international reporting while ensuring these assessments remain relevant for national purposes. ACER UK also emphasises the importance of enabling countries to track progress over time, offering a longitudinal perspective on education outcomes. Ghana and Rwanda are perceived as proof-of-concept examples of how national data can meet global reporting needs without sacrificing local relevance.
- The PAL Network, developed collaboratively in the Global South, is seen as contextually valid and well-positioned to address in-country priorities while contributing to global goals. Countries like Kenya and Senegal highlight the focus on national use of the globally comparable data.

109. These examples demonstrate the *Generate* strand's potential to produce data that bridges global and national objectives, offering both international comparability and insights to inform in-country education action.

110. There are notable limitations to the utility of nationally comparable data for in-country purposes. For example, the PAL Network faces challenges reconciling the needs of member organizations that focus on sub-national comparisons with the global secretariat's emphasis on international reporting. The ICAN-ICARe is designed to be nationally representative. This level of representativeness is very useful for the purposes of advocacy and sensitisation, but it has limitations for other in-country use. For example, it is not yet known whether the data will link to school-based contextual factors to identify potential levers for improving learning outcomes. We found that, in Nepal for example, stakeholders expect data to provide rich insights into community factors, socioeconomics, and other drivers of learning, but resources and sample designs may not support this level of detail.

111. National assessment tools are often more nuanced and better aligned with local context than internationally comparable tools. For example, we observed that praise for the relevance of the work of the PAL Network often reflected previous or other activities carried out by the Network and its members, which are more locally contextualized and would not align with the uniformity required for global comparability and/or reporting.

112. These limitations underscore the tension between designing tools for global reporting and addressing specific in-country needs, which may not always align seamlessly.

113. As discussed in section 4.1, the limited planning for supporting the use of data also risks data not being utilised effectively, and thus risks limiting the likelihood that the *Generate* strand will meet the needs of the *Use* strand comprehensively. Moreover, partners anticipate contextual challenges to connect data tracking and generating to support data use. These include high turnover in government staff, perceived resistance by government to be held accountable for learning levels, and financial constraints. For example, in Ghana, concerns were raised about the affordability and sustainability of a shift to population-based assessment (rather than a sample-based approach).

114. To maximize its impact, then, the *Generate* strand must prioritize early and systematic planning for data use, balance the dual objectives of global and national utility, and address barriers such as government resistance and financial constraints.

EQ5 To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies? (i.e. external coherence at country level)

Finding 12. The D4FL programme activities demonstrate significant efforts to align with national education structures, policies and assessment timelines, and add value rather than duplicate non-D4FL efforts, but challenges remain.

115. There are many examples of partners taking steps to align the logistical aspects of their data generation exercises with national education structures and timelines. For example: The PAL Network is aligning its data collection with the end-of-year school holidays, ensuring minimal disruption to academic activities; in Ghana, ACER UK accelerated initial support to align with the country's 2024 national assessment, ensuring that its contributions were timely and relevant; in Somalia, ACER UK is working to align its first assessment framework with the National Education Framework, which is closely tied to ongoing education reforms; and in Nepal, the World Bank ALMA initiative plans to roll out AMPL alongside the national assessment next year.

116. There is also evidence that D4FL partners are aiming to complement and add value to existing national assessment systems, rather than duplicate efforts. The efforts are particularly in terms of

logistical coordination, capacity building, and policy relevance. The emphasis on policy alignment highlights the potential of D4FL activities to drive systemic improvements in education systems.

- **Aligning citizen-led and other assessments for better data.** The PAL Network emphasises the unique role of citizen-led, household-based, and competency-based assessments in complementing national school exams. For instance, in Kenya, PAL’s work complements national assessments conducted by the Kenya National Examinations Council (KNEC), especially since national assessment data is not publicly available and cannot serve the same accountability purposes. In Senegal, PAL’s individual and household-level assessments align with PASEQ’s school-level focus, creating complementary datasets.
- **Mapping assessments and developing assessment frameworks.** ACER UK aims to avoid duplication by beginning in-country work by mapping existing assessments and assessment systems, ensuring its initiatives align with national priorities. For example, in Somalia, ACER UK collaborates with the education ministry to develop the country’s first national assessment framework, aligning with ongoing education reforms.
- **Building assessment system capacity.** The new grant to NFER aims to build government capacity, enhance expertise, and provide tools to improve foundational learning data processes.⁸ In Ghana, ACER UK is working with national entities to support both international reporting and sustainable national assessment systems. ACER UK’s work in Somalia focuses on aligning the assessment framework with national education reforms, such as teacher recruitment and training. This integration ensures that the framework supports broader system-level changes. The ALMA grants include assessment system capacity building activities.
- **Strengthening policy capabilities.** ACER UK, ALMA and the PAL Network’s *Generate* activities aimed at global reporting include efforts to provide data that can be useful for country policy-making. In some cases, their programmes include activities to facilitate such use. For example, the ALMA programme intends to fund complementary actions (such as analysis and dissemination of data) so that the data from the assessments it supports is used to identify instruction areas requiring intervention, and PAL Network members intend to engage in advocacy so that the evidence they generate influences national education structures and decision-making.

117. These findings and examples point to the important fact that different D4FL partners and strands are setting out to take different approaches to engaging with and supporting government functions. The PAL works outside-in and bottom-up, ACER UK and ALMA provide TA, and NFER supports peer-learning and practitioner networks.

118. This diversity in approaches reflects the flexibility of D4FL’s efforts to reach its ultimate goal. However, challenges remain. For example in contexts like Somalia, where education systems are undergoing significant reforms, the alignment of new frameworks and policies requires ongoing adaptation and engagement. In addition, the effectiveness of alignment efforts will depend on the capacity of national governments to absorb, adapt, and implement the generated evidence and tools. Capacity-building initiatives, such as those supported by NFER, are essential for bridging this gap.

⁸ While this grant was not within scope of the formative evaluation, it reflects a commitment to bolstering national systems.

EQ6 To what extent are D4FL and other FCDO projects complementing one another? (i.e. internal FCDO coherence at country and global levels)

Finding 13. By design D4FL programme works alongside other concurrent FCDO education programmes to improve learning outcomes, but mechanisms to leverage this complementarity in practice are still under-developed.

119. The D4FL is designed to (i) be complementary to other FCDO centrally managed programmes with which it shares the objective of improving foundational learning outcomes, (ii) support and work alongside bilateral FCDO programmes together with the other central programmes and (iii) complement FCDO funding to and influence its global partners like the World Bank, the Global Partnership for Education and Education Cannot Wait which have more financial resources to leverage learning and scale interventions. Together the programmes are aimed at strengthening the global understanding of which reforms work, the global and country availability of data to track progress and monitor the impact of interventions at country level, countries' capability to use data to diagnose learning shortfalls, and testing, adapting and scaling interventions. Key central programmes are:

- The What Works Hub for Global Education (WWHGE 2021-2028), generating evidence at global level on effective education reform; supporting countries using data and evidence to diagnose why children or not learning; and providing support to countries to test and adapt reforms.
- The Scaling Access and Learning in Education (SCALE, 2023-2027), supporting bilateral programmes with expert technical assistance to adapt evidence-based interventions to new contexts and getting them on a pathway to scaling up.

Other relevant programmes are Education Research in Conflict and Protracted Crisis (ERICC), Early Childhood Development at scale (THRIVE) and the Ed Tech Hub. There is evidence in the documentation of the concurrently designed programmes of the effort at coherence and the desire to link up.

120. The evaluation has understood that the D4FL programme management communicates and works with GED managers of the other centrally managed programmes to join up work at the global level, coordinate with global partners, and coordinate support across countries. The evaluation however has yet had limited opportunity to understand better whether, how and how well this works so the programmes' activities leverage one another in practice.

121. At country level, as of now, the evaluation also has very limited evidence of D4FL actively engaging and complementing other FCDO projects in-country. The D4FL activities are still in their initial phases – as is the implementation of complementary FCDO programmes – and opportunities for collaboration may not yet have materialized.

122. Interviews as part of this formative evaluation demonstrate (i) the benefits of FCDO being able to present a "holistic offer" to governments and a consistent view on assessment (including different assessments serving different purposes); (ii) the value that FCDO advisors can add to implementing partners' work where they already "have the ear" of government; and (iii) the potential for D4FL to provide guidance on generating and using outcome data. However, this linking role of FCDO advisors in the D4FL programme is not specified, and currently there is no evidence of D4FL offering such guidance to other FCDO education programmes.

123. At the country level, coordination and active take-up of such opportunities for synergy depend to a significant extent on the FCDO education advisors and implementing partners. In neither case are interest and capability to take up such roles guaranteed. Although the D4FL management actively engages with the advisors, as it does with other central programmes, this is not backed by systemic mechanisms to align initiatives at country level. That said, the evaluation also has no evidence of education advisors not being willing to take up their role: in fact in all the country cases where there were advisors, they were active, as confirmed by advisor interviews, interviews with partners and country authorities. Advisors have acted to avoid duplication (e.g. in Somalia, where the advisor introduced ACER UK to the GPE grants agents, given that the GPE has also mentioned developing an assessment framework).

4.3 Efficiency

124. Efficiency refers to extent to which D4FL is set up to make and making best use of its resources to maximise the chance that programme outcomes will be achieved.

EQ7 Will the programme set-up likely facilitate timely and efficient use of resources?

Finding 14. Differing partner delivery models translate into different timelines, some longer than others. Longer partner timelines to use D4FL resources on the ground are justifiable in terms of relevance or quality.

125. How quickly the D4FL resources are used for country-level track, generate and use activities, depends on partners' delivery models and what they undertook to do.

- ACER UK was contracted in November 2023, and from January to March 2024 it was conducting scoping activities in countries flagged for support. By April it was working directly with government counterparts to provide TA, and aligned its support with country education system timetables. For example, in Ghana it was able to review the draft July 2024 primary 4 and 6 assessments for UIS criteria eligibility in May 2024 already. ACER UK can be on the ground fast, because it delivers the activities itself and does not contract further.
- The PAL Network was contracted in July 2023 to develop and implement globally comparable citizen-led assessments through country-based network members. It moved quickly on setting up a consensus building workshop with experienced network members from 17 countries in the 3rd quarter of 2023; developed the common framework from early 2024 while starting engagement with country statistical agencies; developed quality assurance protocols in the 2nd quarter of 2024; and undertook piloting, tool development and training in the 4th quarter of 2024. Data collection in countries that have agreed sampling frames, will only be in 2025 followed by analysis and engagement on the results.
- The World Bank was contracted in February 2024, and awarded grants to country teams in September 2024. By October country-level implementation had not started. The World Bank process first identified a longlist of countries where support would be strategic, then called for proposals and then adjudicated and selected proposals for funding. Although some country World Bank would have liked grants to be available earlier to align better with country timetables, all countries followed the common semi-competitive process. Activities will be implemented in countries from late 2024 through to 2026.

Finding 15. The D4FL commitment to work with country structures has consequences for the timely and efficient use of resources: partners are leveraging positive factors and taking steps to mitigate negative factors

126. Country circumstances and capacity has already caused delays in deployment of country-level support, and will continue to do so. For example,

- ACER UK's support to Mozambique is suspended because of elections. In the meantime, ACER UK has agreed with the FCDO to provide support to Malawi in response to a request.
- In Ghana, the small teams and limited physical resources like scanners, have led to the Ghana July 2024 national assessment results being delayed for ACER UK's analysis TA to begin. ACER UK has in the meantime started writing the sections of Ghana's report to the UIS that report on the assessment process criteria.
- In Nepal and Ghana the WB ALMA task team lead (TTL) and ACER UK both had to cope with new leadership in their direct counterparts. Both teams have briefed the new leadership.
- In Senegal the Accelerator will contract country-based TA to ensure capacity is available to provide support for EGRA-linked assessments to the capacity-constrained assessment agency.

127. Country circumstances can also support timely and efficient use of D4FL resources. For example in Ghana and Rwanda the country national assessment timetables have created pressure for fast processes, and helped to focus minds.

Finding 16. FCDO's decisions to source partners directly have helped shorten timelines. Ongoing opportunities to maximise efficient use of the programme budget outweigh any potential loss of efficiency due to direct sourcing of partners.

128. The evaluation does not have counterfactuals on what partners' interventions may have cost if they were selected through a competitive FCDO process. In principle, it is possible that partners may have submitted bids that delivered the same services more cheaply.

129. However, across partners resource budgets are sufficiently flexible to facilitate efficiency and there is evidence that partners are intent on using their resource allocations well. For example,

- The PAL Network is considering rules about using unutilized funding in country A to support more expensive evidence collection in B, so that overall use of their grant is efficient.
- ACER UK uses its budget on a case-by-case basis, as approved by the FCDO. Savings in projects can accrue to future projects, but this may create incentives to spend earlier in absence of later demand. However, the buildup of country support cases enables the FCDO to monitor efficiency and country allocations. It is also evident that ACER UK works with small country teams and draws on its internal pool of expertise for specialist experts for as long as needed.
- The World Bank's processes support efficiency because funds would have leveraged existing policy dialogue with government, and were awarded internally to countries on a competitive basis allowing comparisons of the level of activities proposed for the USD300,000 grants. Bank staff is very aware of ALMA and Accelerator resources needing to stretch to fund unfunded TA needs of country programmes.

Finding 17. The D4FL will have to address the gaps in programme design and coherence and up its efforts to deal with barriers to progress, to avoid unproductive use of resources and maximise efficiency.

130. Some evaluation respondents questioned whether supporting many strands towards improved learning outcomes, many pathways within strands, and many countries with sometimes overlapping activities may mean that the programme does not achieve enough with sustainability in any one country, or at the global level to show material progress on learning outcomes by 2027 (see discussion in 4.2 Coherence under Finding 10 for example). Should this be the case, the use of D4FL resources would ultimately be unproductive.

131. On balance, the evaluation team however finds that this risk is unlikely to materialise. There is enough evidence that D4FL will result in additional countries reporting globally, improved national assessment systems in countries, and in demonstration of alternative pathways (see discussion in 4.4 Effectiveness and 4.5 Looking ahead at impact and sustainability, Finding 28). The evaluation agrees with respondents who thought that (i) demonstration of different pathways is valuable in itself, given the tensions around the UIS criteria, countries needing locally relevant, reliable and comparable data, and the number of countries that do not have quality data; and (ii) just supporting *Track*, or *Generate* or *Use* activities would forego demonstrating how all three are needed for impact.

132. However, to minimise any efficiency loss because of the inherent contradictions or tensions in the programme, the programme should address concerns raised under relevance about gaps in its design for data to be used (see Finding 7), about the strategic use of its resources (see Finding 8), and coherence (Finding 10 and 104), and address barriers to progress (see Finding 25).

EQ8. To what extent are partners responding in a timely way to requests for support?

Finding 18. Partners that respond directly to country requests, have moved swiftly to provide support.

133. ACER UK is the only implementing partner that mobilises resources to respond directly to countries' demand for support as they occur. ACER UK has been swift in responding to needs, aligning its support with country education assessment timelines, as confirmed by ACER UK reports, and government stakeholders in the Ghana and Somalia case studies. In Ghana government has noted that ACER UK is pro-active in agreeing support with it, and the value of this in enabling it to maximise its chances of reporting sooner. ACER UK has also responded fast to delays outside of its control, e.g. shifting from Mozambique to Malawi.

134. The World Bank FLC programmes respond indirectly to country needs as articulated by the World Bank country teams in their proposals for support, in the context of World Bank country programmes and dialogue. As the country teams have not yet started implementing the grants, the evaluation does not have evidence on the timely agreement at country level on activities to be implemented, following the grant award. There is evidence, however, that the programme's delivery model can mean some misalignment between money reaching the ground for implementation, and country assessment timelines, because there is a common process across countries for grant awards which may or may not be aligned with the timing of individual countries' needs. As the work continues and future funding rounds become available, the World Bank should clearly communicate timelines to country teams so that they can plan accordingly.

135. The PAL Network does not respond to requests under the D4FL funding, but rather selected Network members for implementing ICAN/ICARe based on their experience in implementing learning assessments. Processes to agree sampling frames with government statistical agencies are delayed, especially where Network members have weaker existing relationships with government.

EQ9 To what extent have D4FL resources been used efficiently to date, and its intended timeline kept?

Finding 19. Early evidence is that partners manage processes well, deliver as planned, and that there are efficiency-positive factors in the delivery models of each partner.

136. Initial evidence is that partners manage processes well, and have delivered as planned, with relatively minor deviations that can be justified, as detailed in Table 6 below.

Table 6 Evidence of the delivery timeliness and efficiency to date

Assessment criterion	Selected examples from the D4FL programme to date
Partners' work has been well project managed to result in quality outputs being delivered within expected level of effort	Across partners, the interviews conducted at global and country level did not raise any alarms on the quality of partners' project management, despite delays. There is some positive evidence: counterparts have confirmed that ACER UK's support in countries is well-managed; the World Bank is backing country education teams with management resources at regional and global level. How well this will translate into country-level implementation is yet to be evaluated.
Partners' work has involved minimal waste and duplication of efforts	The PAL Network's proposal set out how it will maximise efficiency and ensure countries selected to go forward are those where there is groundwork has been established, and team structures are in place. The relatively small World Bank FLC grants leverage many-fold more World Bank resources to improve assessments and learning and leverages existing World Bank relationships with country authorities. The D4FL leverages ACER UK's familiarity with global reporting criteria, and assessment expertise, which has meant fast responses in Ghana and Rwanda.
The programme timeline is being kept. If the programme timeline is not kept, the deviations are merited and likely to support better achievement of outcomes.	The PAL Network has been on-time with developing its globally comparable assessment framework, including processes, documentation, frameworks, manuals and training data. The Network still believes it is on track to collect data in 2025, despite slow responses from some countries' statistical authorities. ACER UK has been on-time with delivering its outputs where it has controlled processes, e.g. in Ghana and Somalia (confirmed by authorities), and according to its reports, in Rwanda and Malawi. Timelines slippages have been due to factors outside of its control (e.g. two-month delay in Ghana due to assessment data processes taking longer than expected). The World Bank's timeline for sending out, receiving and awarding contracts have been slower than planned (completing in September rather than July 2024), but will enable it to start supporting countries' processes from end 2024 to more countries than initially proposed. No evidence yet on implementation timeliness.

137. Some global stakeholders have challenged the efficiency of undertaking PAL Network assessments in countries that already report or where the state is already prioritising international reporting and national assessments. This is discussed under 104 paragraph 105.

Finding 20. The FCDO’s management of the programme is robust relative to the management resources dedicated, and partners are satisfied with the support provided by the FCDO to keep work moving at pace

138. The FCDO team is relatively small (resources equal 0.8 of a full-time staff member from two staff members, and support from the GED evaluation expert up to December 2024) and undertakes both oversight of implementing partners and engagement with partners and other stakeholders at global and country level towards programme performance. In the judgement of the evaluation team, if the latter responsibilities are to receive more attention, the team would need to grow.

139. Even so, partners and stakeholders were uniformly complimentary of the role played by the FCDO at global level. Partners thought FCDO is providing the right level of support to keep work progressing. Partners generally agreed that better coordination would benefit the programme and their own work. However, some partners are concerned about the level of cooperation that may be required and whether they can deliver given different positioning in countries’ institutional setup.

4.4 Effectiveness

140. Effectiveness refers to the extent to which D4FL is likely to achieve intended results.

EQ10 To what extent are D4FL activities aligned with the programme ToC?

Finding 21. There are some outcomes in the ToC that is not well-supported by D4FL activities, and some activities of the D4FL that are not well articulated in the ToC even although they support the articulated programme impact or outcomes.

141. The question on the alignment between actual and planned D4FL activities and the programme ToC has been covered extensively under the Section 4.1, Finding 4, Finding 5 to Finding 7. Without repeating the salient detailed evidence, key aspects are:

D4FL activities supporting D4FL outcomes

- *Activities towards Outcome 1:* UNESCO UIS, GEM-R, the PAL Network, and the World Bank ALMA activities are oriented to Output 1 (Fund generation, coordination and publication of comparable learning data using quality assessments) and 2 (Fund development and implementation of global methodologies to track foundational learning for all internationally), which in turn is expected to contribute to Outcome 1 (More LICs and MICs report internationally using comparable and disaggregated data on learning, including in emergency contexts.) See detail in Finding 4.
- *Activities towards Outcome 1 & 2:* World Bank ALMA and at least one Accelerator grant, as well as ACER UK’s TA in practice to date, are supporting Output 3 (Respond to national opportunities to adapt and implement rigorous regular primary learning assessments), which contributes both to Outcome 1 and Outcome 2 (More state capability to regularly use disaggregated learning data so more governments and teachers implement and monitor interventions to improve foundational learning for all in priority countries.) See detail in Finding 5.

D4FL ToC aspects that are not well supported by D4FL actual and planned activities.

- *The ‘in emergency contexts’ aspects of Outcome 1 and Output 1:* The D4FL is implementing activities in emergency contexts (e.g. Somalia) that may support achievement of Output 1 and

Outcome, and although partners are expected to think about integrating assessment in emergency context into their work, this is not yet a systematic part of their work. This is discussed under Finding 5.

- *Use of learning data for decision-making (Outcome 2):* ACER UK, the World Bank ALMA and the PAL Network intend to support countries to or to produce data that can be used nationally, and promote the use of data in countries, but there are no specific plans to support countries directly to strengthen their capability to use learning data.
- *Use of assessment in classrooms (Output 5):* Whereas the PAL Network contributes to communities using learning assessments, there are no activities to support the use of assessments (or assessment data) in classrooms.
- *Inclusion of children with disability in data generation for global reporting and country use:* The evaluation has found limited activities to ensure that children with disabilities, including visual and hearing impairment, are included in national assessments. In Ghana ALMA plans to support primary level 2 assessment in 2026 to include assessment of children with disabilities.

D4FL activities that are not well articulated in the ToC

- *Accelerator Activities:* The Accelerator provides country grants to strengthen the instructional core and related delivery systems, reinforces regional technical expertise on foundational learning and supports foundational learning knowledge creation and exchange. While the programme therefore contributes to more state capability to implement learning interventions, there is not an explicit and clear link between learning data and Accelerator activities, as articulated in TOC outcome 2 as the programme objective, even if in some grants activities may have been identified based on learning data or Accelerator activities may lead to greater demand for learning data (see paragraph 83 in the discussion of Finding 7).
- *Disaggregation for global reporting, versus disaggregation for national use (Outcome 1 and Outcome 2).* By stating Outcome 1 and 2 as equal aims, the ToC does not recognise the tension between disaggregated for global reporting (by sex, reachable through a representative sample) and for national use to improve curriculum and instruction, where countries have specific disaggregation needs to be met (i.e. between districts or schools to identify districts that need support), or disaggregations that cannot be applied (i.e. refugees), so that standardisation is difficult and different samples are required.

142. In short, the review of D4FL partner documentation, interviews with partners and country case studies did not reveal D4FL activities that are not likely to contribute to programme impact or outcome. But the ToC as it stood at the start of the formative evaluation did not articulate all pathways that are present in practice clearly. In addition, not all the outcomes and outputs are well supported by already contracted activities. Programme activities and investment is weighted towards *Track*, which may leave the programme short on demonstrating tangible results for *Use*.

Finding 22. There is mixed evidence on the extent to which plans and activities under the *Generate* and *Use* strands are aligned with the D4FL guiding principles.

143. The D4FL business case set out guiding principles for improving the *Generation* and *Use* of learning data. While some principles are operationalised, others are less so.

Table 7 Alignment between D4FL guiding principles and *Generate* and *Use* activities

Principle	Adherence of D4FL activities to principles
Nationally and sub-nationally owned; and answers problems faced by countries, and effectively responds to their own demands and targets.	<p>The ACER UK activities tend to be nationally owned, because the programmes are in response to demand articulated by government agencies (e.g. Ghana and Somalia)</p> <p>The data that the PAL Network generates is owned by the Network members, but not necessarily by national and subnational governments. The Network believes that involving government stakeholders from the start will increase buy-in and trust in the data, but ownership and willingness of Government to use the data is untested.</p> <p>The World Bank ALMA <i>Generate</i> and Accelerator <i>Use</i> activities should be nationally owned, given that it supports World Bank operations implemented through governments, but the evaluation has contradictory evidence from the small sample. In some countries, government buy-in may not be universal, as governments are not homogenous and do not always agree internally. However, there is also evidence of countries where the evaluation could confirm government buy-in.</p> <p>The NFER / KNEC grant is aimed at support that is nationally demanded and owned.</p>
Includes teachers and schools leaders and supports them	<p>There is no evidence found in the formative evaluation data collection, that use or generate activities included teachers and school leaders, and will support them to effectively do their job. The Accelerator programme includes activities aimed at teachers and school leaders (e.g. teacher support and professional development)</p>
Cost-effective and sustainable processes (i.e. not too burdensome for governments, schools and teachers), but allows for rigour, quality and comparability.	<p>There is mixed evidence on the cost-effectiveness and sustainability of processes. TA through ALMA and ACER UK to strengthen national assessments, should be cost-effective as they leverage assessment paid from somewhere else, and the TA includes capacity buildings. But, these assessments may not be sustainable. In Ghana, for example, government noted that the results of the ACER UK support will be sustained if government continues to receive support for undertaking assessments.</p> <p>The sustainability of the PAL Network work, which is outside of government systems and dependent on donor funding, was also questioned. Ironically, at the same time many respondents saw the Network as essential to sustain demand for data.</p>
Effective at building state capability (not reliant on external or parallel processes).	<p>In all the country cases from ACER UK and World Bank, support has been or will be provided by working closely with government institutions. Besides planned workshops to build the capacity of state agencies (NaCCA in Ghana, QASD in Somalia) the interventions work with government institutions (triangulated for ACER UK UK, but not World Bank as activities have not started).</p>
Influences wider education spending, policies, and programmes	<p>There is limited evidence that D4FL <i>Generate</i> activities aim to or are likely to influence wider education spending, policies and programmes. Promising factors appear in some country cases (e.g. Ghana), such as policy makers working with assessment authorities to manage the support. ALMA includes analysis of results to stimulate use in some grants. But, at the same time, there is limited attention yet, and now evidence on whether support for generating data will translate into using data in these ways. The PAL Network intends to engage stakeholders across sectors of society and at all levels using their data to advocate and pressure for change.</p>

EQ11 What has the programme achieved to date?

Finding 23. Partners are making expected progress in delivering activities and outputs, and early results indicate that there will be at least incremental achievement on the programme outcomes within the programme timeframe.

144. This finding is evidenced in paragraph 136 and Table 6 on delivering activities as planned. With regards to outputs contracted, early indications are that outputs are being delivered at country level, to deliver on the programme outcomes. The evaluation finds that more LICs and MICs are likely to have reported internationally, using comparable and disaggregated data on learning, including in emergency contexts (outcome 1), by filling global learning data gaps, and linking existing learning data. In this regard the country case studies, document review of partners' reports, and interviews with partners and the FCDO globally and at country level,⁹ indicate that it would have:

- **Funded the generation, coordination and publication of comparable learning data (Output 1):** Other than the work by GEM-R and UIS (not evaluated), ACER UK has, and ALMA will be supporting countries generating comparable disaggregated learning data up to UIS quality standards. These assessments will be focused on in-school children to generate comparable learning data, and will include emergency contexts, at a minimum Somalia where ALMA will support a first comparable national assessment. In some cases, children with disabilities will be included, but current evidence on this is limited to one case. While these activities have begun, they have not yet produced data as collection and/or analysis processes are yet to happen, and acceptance by the UIS is not guaranteed. However, the evaluation has found no reasons to doubt that the activities will result in more generation of comparable data.
- **Funded development and implementation of global methodologies (Output 2):** Other than the work of UIS to develop the AMPL, pilot it and implement it in more countries (not evaluated), the PAL Network's ICAN/ICARe assessments are relevant. The development is completed for the most part, and the Network says it is on track to implement the methodology in up to 13 countries in 2025 (see Table 4 in section 3.2 on programme implementation), with the assessments unlikely to go ahead in two more countries, because of the cost of data collection. The PAL network assessments will collect data also for out-of-school learners, and will deliver disaggregated data by sex. While children with disabilities have not been included to date, the PAL Network has said that it is working on including these learners (although not under the D4FL funding). ACER UK is contracted to provide a global public good on aligning national learning assessments with global reporting by December 2024, but the UIS criteria was only published on the 9th of December 2024.
- **Responded to national opportunities to adapt and implement rigorous regular primary learning assessments (Output 3):** ACER UK is making good progress on delivering on its contracted output targets (that respond to national opportunities) on time (2 countries where reporting is ensured by December 24). It is active in four countries, but in two the initial work is around developing a framework on national assessments, which should also contribute to the output (by contributing to better choices between which assessments are done by whom for which purposes). Besides adapting the assessments for global reporting in Ghana, it has helped improve the quality of the assessments for national purposes, e.g. it helped to link the 2024 to 2022 assessments and will help the government to interpret achieved coverage for representivity. The World Bank ALMA programme has made progress in supporting 12 countries (11 under ALMA and one under Accelerator) to adapt and implement primary learning assessments. Implementation is starting from October 2024. This is more

⁹ To note, country level FCDO interviews were not possible for Ghana, and triangulation of partner and FCDO information with government counterparts were not possible in Senegal and Nepal in the data collection window.

countries that what was committed in the D4FL results framework, and additional grants may be made in 2025 (depending on release of further funding).

145. More state capability to regularly use disaggregated learning data so more governments and teachers implement and monitor interventions (Outcome 2) can be expected to increase by the end of the programme, at least for some countries in principle. However, progress on the sub-components of the outcome (more state capability to regularly use disaggregated data; and more governments and teachers implementing and monitoring interventions), is not necessarily linked.

146. It is likely that countries will have improved capacity to generate data, given *Generate* interventions. There are also efforts to make the data usable, which would be component of 'capability to use' and may lead to actual use in certain contexts, although the evaluation currently has no evidence of this happening. At the same time more interventions to improve foundational learning is likely, because of Accelerator, but the programme does not directly support the utilisation of learning data for these interventions.

147. Evaluation evidence (see sources in paragraph 144) indicates there is some progress on the outputs linked to this outcome, which will result in some change at country level. By output, with early results detailed, by the end of the programme, D4FL would have:

- **Embedded learning assessment in national policies and reforms on curriculum and teaching:** ACER UK's work on National Assessment Frameworks in Somalia (linked to the Education Sector Policy) and Malawi is likely to deliver results in 2025. There is work contracted under ALMA for similar support to countries (e.g. Nepal), as well as support in establishing permanent mechanisms to disseminate data as part of education sector systems. The evaluation did not learn, however, of any other specific efforts towards this output.

The Accelerator Programme, the closest fit to embedding assessments in reforms because it supports reforms on curriculum and teaching, has made progress in setting up country-level TA in 12 countries (two more than committed to in the D4FL results framework)¹⁰, but the role of assessment data and evaluation in the interventions is explicit only three of the countries. The FLC proposal for D4FL funding says that ALMA generates data, and the Accelerator uses the data, but they overlap in only four supported countries, and were awarded at the same time.

- **Strengthened use of simple foundational learning assessment in community interventions, but not in the classroom:** The PAL Network's work on citizen-based comparative assessment is at country-level about members leveraging the data to promote education interventions. There is not significant progress on the use of foundational learning assessment in classroom interventions as such, because activities to support this link directly (i.e. to ensure that quality assessment for learning is applied in interventions to support instruction, or in classroom teaching) are not present. Accelerator activities exist to improve instruction but are not strongly linked to assessments, and there are activities aimed at making data from D4FL-supported assessments usable, but as argued in Finding 7, this is necessary but insufficient to result in actual use of data for interventions.

148. Although not linked in the ToC to a distinct capacity building output, current evidence is that the programme is likely to have strengthened capacities in national assessment systems. Capacity strengthening activities are key components of the D4FL programme in all countries and government

¹⁰ To note that another output in the results framework towards results from Accelerator is lagging against the milestones set out in the D4FL proposal, namely imbedding expert capacity with global and regional teams. FLC notes that they are working to identify regional needs and best modalities for support.

at different levels is expected to have increased data awareness and to be better able and willing to use data to address and improve foundational learning by the end of the project. All the country cases have illustrative examples: ALMA in Somalia will train the National Assessment Team, ACER UK's work on the National Assessment Framework in Somalia also includes technical support and capacity building of the team; in Senegal, the Accelerator programme is funding a technical expert to provide TA in person to contribute to sustainable capacity in country; in Ghana technical support is provided by both ACER UK and ALMA to the NaCCA; and in Nepal the ALMA grant will support training at national and provincial level on assessments, as well as the training of non-government bodies to mitigate against trained government staff rotating out of relevant positions.

Finding 24. From the perspective of experimenting with different options to generate data for SDG 4.1.1, there are early promising results.

149. Before the start of the evaluation, FCDO had already supported the UIS pilot application of AMPL in 5 countries (Lesotho, the Gambia, Kenya, Nepal and India). The ACER UK support in Ghana, and Rwanda¹¹ to adapt national assessments to be eligible for global reporting under the UIS criteria, and the ALMA grants – some of which imbed AMPL in (Ghana) and other that run AMPL alongside national assessments (Somalia and Nepal) – are further examples of likely successful experimentation. Using a different approach, the PAL Network has developed robust tools to report. That the tool is robust was confirmed by global stakeholders and partners. If it is implemented successfully, some countries may agree to the data being used for global reporting, but this is not guaranteed. Nor is the UIS acceptance of the results. Through these various pieces of work the D4FL will most likely successfully demonstrate that different approaches can be used (in complementary ways) to report. Programme stakeholders hope that this will make SDG4.1.1a more likely to remain as a globally tracked indicator, as well as increasing the number of countries that have reported.

150. As noted in the relevance discussion (see Finding 8), it is less clear that in aggregate the D4FL will result in scalable models of generating globally comparable data applicable beyond the countries supported by the specific partners. It is not clear that actors other than the selected partners' approaches will be seen as providing eligible approaches, that the selected partners can support on a global scale, or that D4FL country selections will by 2026 provide examples of both countries that are well placed to undertake the different models (proving that a model can work), and countries that are not well placed (proving that it can work anywhere).

EQ12 What are the likely barriers and facilitators to achieving results?

Finding 25. Political will, entrenchment of assessments in policies and laws, country capacity, assessment relevance, and how assessments are used, help or hinder whether countries *Track* learning, and *Generate* and/or *Use* learning assessment data.

151. The presence or absence of these factors were highlighted across correspondents as helping or hindering countries reporting internationally, generating regular comparable disaggregated data nationally, and using data to improve curriculum and instruction. For example:

- As noted in the Relevance discussion, political will to generate learning data and track learning nationally and globally, can falter if comparisons show poor or worsening learning. When

¹¹ In Rwanda, the UIS is at the same time working with ACER UK's GEM team (not funded by D4FL) to implement a AMPL alongside the work on imbedding UIS criteria in national assessments.

political or technical government and institutional leadership changes, support for learning assessments can also shift. (See also discussion in paragraph 67 under Finding 3).

- In many countries learning assessments have not been entrenched in education policies or legal frameworks, making it easier to stop regular assessments and default to tracking enrolment and progression indicators. It can also result in overlapping assessments being undertaken, none of which are fully aligned to country needs, because these are not articulated.
- Country capacity (assessment systems, human resources, funding, equipment) is a major barrier to generating regular comparable data nationally, and comparable data to report globally. It is one of the main reasons why there are trade-offs between the *Track* and *Generate* objectives of the D4FL. Amongst the country cases there are examples where capable and capacitated country counterparts facilitated D4FL activities, but also the opposite. As noted in Relevance Finding 7 paragraph 85, usable data is a first necessary step to using data, but on its own is insufficient to translate into interventions and reforms on curriculum and instruction. This requires additional capacity, which is often lacking.
- Whether assessments are relevant to countries needs and curricula can be a barrier (if not) or a help (if they are). As noted under Relevance (Paragraph 67), country authorities can see international reporting requirements as imposed and not relevant for the country, especially if assessment does not align with the national curriculum. If there are too many protocols before data can be released, data can be too late to be relevant for shifting country learning practices.
- How the purpose of assessments is framed, influences willingness to conduct them and use the data in country. When assessments are framed predominantly as being about diagnosing shortcomings in policies and delivery of foundational learning to support stakeholders, they are more likely to become imbedded, than when they are predominantly about (political, institutional, teacher) accountability for learning. Stakeholders also noted that when assessments and interpreting their results are highly technical, the results can be obscure for those who need feedback or have influence.

Finding 26. Stakeholders also noted that changing criteria for UIS eligibility and tension about the extent of standardisation that is needed for global comparability, are barriers specifically to generating globally comparable data.

152. Independent experts and programme stakeholders noted that the bar on what would be considered eligible has shifted over time. Initial plans were to monitor foundational learning against countries' own curriculum expectations. Later policy linking was introduced in consideration of varying expectations and varying assessment types and levels of difficulty. Recently this has shifted to standardising assessments themselves across countries using common assessments and common items. This is not conducive to different actors, including country governments, making the effort to link assessments to generate national comparable data to eligibility to report globally.

4.5 Looking ahead to impact and sustainability

153. The evaluation framework required the articulation of pathways and opportunities for sustained impact and influence to inform the summative evaluation. Because of the nature of the questions in this section, it is forward looking, blurring the line between findings and recommendations.

EQ13 What are the likely pathways and timelines for outcomes and desired impact?

Finding 27. Systematic achievement of ToC outcomes and impact is unlikely within the timeframe of the programme, because country assessment and reform cycles take time. However, the formative evaluation provides good pointers to pathways and criteria that can be used to evaluate progress to outcomes and impact in the summative evaluation.

154. There is broad consensus among stakeholders that the planned outcomes and impact of the D4FL programme will take time to achieve, though timelines towards improved practices and better learning outcomes, will vary depending on where specific countries supported by D4FL are at, and whether and how quickly other countries/actors may adopt successful demonstration of pathways. For example, Somalia is at the start of its journey, currently developing a Learning Assessment Framework, at the same time as going through significant reforms of its entire education system in what is a very complex (emergency) context, and in a next step is planning to conduct a national learning assessment with support from the World Bank ALMA and report internationally for the first time. Other countries such as Ghana has already developed a Learning Assessment Framework, and is conducting national assessments. The D4FL work is aiming to help Ghana use these assessments to report internationally, and supports interventions on curriculum and teaching. It may therefore have progressed further by 2027.

155. There are however early signals of progress towards outcome and impact-level results that can be evaluated in the summative evaluation if set out clearly in the ToC. The findings of the evaluation under Relevance and Effectiveness are germane.¹² The ToC as formulated by the D4FL stands for the summative evaluation to a significant extent, but some changes and additional clarity are required. These are detailed in the recommendations.

156. In addition, because of the trade-offs between generating data for global reporting versus for country use and to support assessment in the summative evaluation, there should be greater clarity on which of these outcomes would be prioritised. This is not reflected in the amended ToC.

157. There are several indicators associated with the theorised ToC result chain that can be used in the formative evaluation framework (as assessment criteria) to indicate the likelihood of outcomes and impact. These span the full ToC, and will therefore allow evaluation of progress from inputs to outputs, to intermediate outcomes and outcomes. Annex 8 provides a list of potential assessment criteria. The selection of indicators and their phrasing will be refined in the summative evaluation framework, due in February 2025 after engagement with the FCDO and partners.

EQ14 What are the avenues for D4FL to have sustainable impact?

Finding 28. Partners' plans include some conceptualisation and integration of avenues to sustainable results, and a good understanding of the barriers, but sustainability can be integrated better.

158. Avenues to sustainable results as conceptualised by the FCDO and partners are:

¹² Sections 4.1 on Relevance (Finding 4, Finding 5 and Finding 7) and 4.4 on Effectiveness (Finding 21) discuss the validity and comprehensiveness of the ToC in setting out the pathways. Finding 23 and Finding 24 in the Effectiveness section also discuss which pathways are likely to ensue, and Finding 25 discusses the barriers and facilitators of achieving the pathways.

- *Strengthening capacity of relevant government in-country to be able to implement – or outsource and manage – learning assessments.* This can be done remotely or in-person at Ministry level for a sufficiently long period to ensure capacities long term. This is for example integrated into programme design through ACER UK and the World Bank in Somalia and Ghana, and in Senegal through the Accelerator programme.
- *National dissemination of results from supported assessments to stimulate demand for learning improvement.* If data is not analysed and results shared, then the risk of data not being used is higher. This is a focus of the PAL Network. It is also a focus of ALMA grants, e.g. in Ghana.
- *Awareness-raising activities at community and supra-national level.* The PAL Network’s plans conceptualise community-level engagement as well as engagement with regional networks as an avenue to support bottom-up demand for government accountability to use generated data to improve foundational learning.
- *Supporting government buy-in at all relevant levels from the start of the intervention.* This is seen by partners as needed to create the best possible conditions for sustainable impact. This is especially important with devolved governments where responsibility for primary education does not lie with central government, for example in Nepal or Somalia.
- *Conceptualising D4FL-funded activities as one piece in a larger endeavour or programme of work.* The World Bank and the PAL Network do not conceptualise their D4FL activities as standalone. For the World Bank D4FL activities occur in the context of ongoing engagement with countries which may support continuation of D4FL-linked results in future. For the PAL Network, it is likely that supported partner organisations will continue their membership of the Network and engagement in learning assessments and outcomes, which may similarly sustain results in future.
- *Producing global public goods.* The D4FL includes funding to produce global public goods, such as tools and guidance materials and knowledge creation on foundational learning, to enable countries to implement assessments independently.

159. As much as capacity and (political) buy-in from all levels of government are facilitators when they are in place, they are also significant barriers to countries continuing to undertake assessments and using these assessments to improve instruction and curriculum. Partners have significant activities on capacity building for assessment, but as discussed under Relevance, Coherence and Effectiveness, insufficient attention to how they engage with policy makers in these activities and through additional activities to create an enabling environment for use.

160. The lack of activities supporting the *Use* objectives of the D4FL programme is critical also for sustainability. If demand for learning data to shape and monitor countries’ reforms is not institutionalised, it is more likely that reluctance to undertake assessments because they may demonstrate a lack of learning progress will prevail.

EQ15 What are the avenues for FCDO to influence the wider sector?

Finding 29. The D4FL benefits from the FCDO’s overall strategy to influence the wider sector on learning outcomes, but can be more explicit on how it will influence its own partners and leverage its results for the wider sector.

161. D4FL, as one of several FCDO programmes to improve learning outcomes, shares in the GED’s overall strategy to influence the global education community on the importance of learning data, and

using data for evidence-based reforms to improve teaching. This includes, as is made explicit in the business case, leveraging FCDO representation in the governance structures of global partners funded by the D4FL and other FCDO programmes, to drive programme delivery.

162. While the evaluation team learned of efforts by the FCDO to engage its different partners on programme direction, the extent to which it can do this differs. While the FCDO can steer the work of ACER UK and the PAL Network (managed as accountable grants), the World Bank and UNESCO funding are multi-bi contributions (earmarked bilateral funding for multilateral organisations) managed under memorandums of understanding and with less direct say. Better internal coherence requires a strategy to bring the different strands and partners together.

163. It is also not clear to the evaluation exactly how the FCDO expects the D4FL's results to be leveraged beyond the programme itself, other than through global public goods produced by some partners. These may capture some results for global use, but on their own will not lead to non-D4FL countries and global stakeholders prioritising quality, disaggregated, comparable learning assessments and data, adopting D4FL-supported tools and methodologies, and using generated data for evidence-informed policy making.

Finding 30. The D4FL programme provides the FCDO with strong opportunities to influence the wider sector using the knowledge generated from its activities.

164. At the global and country levels, FCDO is well-positioned to advocate for and influence the broader education sector. The evaluation learnt that the FCDO is widely regarded by partners as a relatively neutral actor on learning assessments and monitoring learning, allowing it to serve as a trusted collaborator, fostering partnerships and meaningful sector-wide dialogue.

165. The D4FL is likely to deliver significant learning, given its experimental nature. There is an opportunity for the D4FL to surface this learning across partners, countries and interventions, for sharing at global and regional level. This should include any proof of concept that emerges, but also what has not worked and how interventions can address different context barriers.

166. At country level, the FCDO should consider how it can influence sustainability and use of assessments by ensuring, by modelling the importance of using data in decision-making and demonstrating early wins on the value of quality learning assessments for progress on learning. There is some of this already built into the ALMA programme, and understanding what works under which circumstances will be important. There are already good examples from the country cases of the role the FCDO can play, for example in Nepal, where it is working to facilitate dialogue between implementing partners and government departments to ensure that the importance of learning data and assessments remains at the forefront of education policy and programming. Success in improving foundational learning at the national level could also serve as a model for neighbouring countries, encouraging broader regional adoption of similar reforms.

5. Conclusions and Recommendations

5.1 Conclusions

167. The TOR posed two overarching questions for the evaluation. This concluding section will answer the first, namely: *Is each aspect of the programme being implemented well, and working together to lead to the intended programme outcomes?* It will do so by first discussing conclusions against each of the evaluation criteria, before providing a concluding answer.

Formative evaluation conclusions

168. This formative evaluation of the D4FL programme occurs early in programme implementation. While partners under the *Track, Generate* and *Use* strands of the programme have been contracted, programme implementation to date is largely at the global level to develop learning assessment instruments and decide in which countries learning assessments and interventions to improve the instructional core of foundational learning will be supported. Country-level implementation to date has been the UIS AMPL pilots (outside the scope of the evaluation except for coherence), the PAL Network's preparation for citizen-led assessments to generate globally comparable data, and ACER UK's support to Ghana and Rwanda to adapt their national assessments to be eligible for global reporting under the UIS criteria, and support Somalia and Malawi to develop Learning Assessment Frameworks. For this reason, the evaluation report has put significant emphasis on relevance and coherence of the programme, and looked to learn early lessons on likely efficiency and effectiveness.

169. **Relevance:** The D4FL programme comprises objectives and activities that are relevant to the needs of LICs and MICs and global policy makers and stakeholders in foundational learning outcomes and assessments. Its end goal – to improve foundational learning outcomes – is highly relevant in a context where learning progress is slowing. The D4FL ambition to increase the availability of disaggregated learning data that can be used to design and monitor foundational learning interventions is germane to achieving improved learning outcomes, as is the objective to increase state capability to use foundational learning data. The objective to fill global learning data gaps through more LIC/LMICs reporting internationally is relevant to fulfil the SDG commitments and align with current monitoring systems. There are ongoing tensions in the global dialogue on how much the sector should invest in global monitoring and reporting, given potential trade-offs with temporally comparable quality learning data generated to meet countries' specific needs and waning momentum behind international reporting.

170. **The D4FL programme exhibits this tension, but does not resolve it in its design.** This may hinder the programme achieving its targeted impact (progress on learning). The programme activities are significantly oriented towards international comparable reporting and tracking, with the funding for UNESCO UIS, GEM-R, the PAL Network, ACER UK and the ALMA initiative working towards more countries reporting quality data globally. The D4FL deliberately funds different ways to achieve this goal, to demonstrate that globally comparable data can be generated in ways that are relevant to countries. The ACER UK, ALMA (to some extent) and PAL Network initiatives aim to work within or alongside country assessments, or contextualise globally comparable instruments for countries. Such alignment is relevant for country and global purposes, and could mitigate countries' resistance to implement global assessments they believe are not aligned to country context. However, the balance of funding and activities for *Generate* is oriented towards *Track* rather than *Use*. This implications for programme effectiveness, including on its equity objectives, impact and sustainability.

171. There are also gaps in the extent to which the activities will deliver the programmes' objectives. These include the extent to which the programme activities will leverage specific opportunities for or address specific barriers to international reporting of learning data in emergency contexts. While partners may work incidentally in countries that are in or affected by emergencies, with moving populations of refugees or internally displaced persons and other challenges in such environments such as the variety of actors providing education services and security challenges, there are no activities funded by D4FL to achieve the programme's stated goals on EiEPC.

172. Another gap is on increasing state capabilities to use such data. While the evaluation found evidence of partners supporting or thinking about what is needed from learning data and how it should be disseminated so that it can support countries' use of it, there is not a substantive stream of work to build states' capability to do so. Data on its own is insufficient, even if designed to be usable. In addition, on usability, the evaluation found that not all partners sufficiently consider how data will be used up front, but see use as something that comes after the data is collected and analysed. A specific gap is on strengthening the use of simple foundational learning assessment in the classroom.

173. While the World Bank Accelerator programme clearly contributes to the programme's ultimate aim, its connection in practice with foundational learning data in most of the grants, is not direct assistance to build countries' capacities to use learning data for the design of its interventions. Rather, the link to learning data is more indirect, namely through the expectation that this design would have been guided by learning data, and may trigger more demand for learning data.

174. The processes for deciding who should support whom, and how, under the D4FL initiative are effective in meeting immediate demand and leveraging partner expertise, but may not be fully strategic against the overall D4FL objectives. The FCDO selected the partner to support based on their credibility, reputation and experience, and as processes to select countries are partner-specific, and demand-driven (except for the PAL Network, which is purely on where it has experienced partners) there is generally a match between partner expertise and countries' needs. However, because it is demand-driven and not coordinated across partners, the net effect does not equal the programme optimally meeting the most urgent LIC/MIC needs across countries and may not deliver proof of concept for FCDO on global reporting.

175. Partners' processes for identifying and ensuring that TA meets the needs of the selected countries, demonstrate promising early results, with evidence of tailored scoping, contextualization, and strong partnerships. But contextualisation in D4FL-supported efforts to collect globally comparable data, can be limited.

176. In summary, although the D4FL's objectives, design and activities to date respond to and meet the needs of LIC/LMIC countries and contribute to global needs of the sector, there are some gaps between programme design and objectives that can be addressed, specifically on how the programme balances data generation for global reporting versus country use, EiEPC, use of data and the extent to which the selection of countries to support is optimal for programme purposes across partners.

177. **Coherence:** Partners share a unified aim of improving learning outcomes, agree on the importance of quality foundational learning data and agree on what is meant by quality data, but across partners the programme is not fully coherent. Firstly, the *Track* strand incorporates tensions due to uncertainties around global certification requirements (UIS sets the criteria and certifies which data will count, and the programme is experimenting with different ways of delivering globally comparable data, not all of which may be eligible). Secondly, partners anticipate challenges in connecting data tracking and generation, to supporting data use. It is not that there are contradictions between what is done in the *Use* strand and *Track* and *Generate* strands, just that not

much is done in the *Use* strand so that the different components do not pull foundational learning in equal measure, towards better outcomes. Thirdly, at present interactions between partners have been minimal, reducing opportunities for alignment and integration and resulting in potential for actual or perceived duplication.

178. Externally, the efforts by partners to align the logistical features of data generation exercises with national education structures and assessment timelines, means that generally the programme is compatible with in-country education and assessment systems, structures and policies. There is also evidence of the programme contributing to reduce overlap and duplication between learning assessments at country level, with partners working on helping countries to develop Learning Assessment Frameworks, that can identify gaps and guide countries' decision-making on which assessments to allow and help them coordinate between partners.

179. **Efficiency:** There is an obvious consequence of the tension between different programme activities in the *Track* strand, and whether the programme set-up will facilitate using resources optimally to achieve programme objectives. This is linked to the question as to whether the programme is trying to do too much, and therefore end up not doing anything well enough to move the needle on its main objectives, with the result that the D4FL resources are, ultimately unproductive. The evaluation concludes that this is not the case, provided that issues about design, coherence and the barriers to progress are addressed.

180. At the level of the activities being undertaken, however, early evidence shows efficiency. While the different delivery models of partners translate into different timelines, some with longer time lapses before activities on the ground in countries commence from when D4FL resources are available to them, these timelines are individually justifiable by the extent to which longer preparation time contributes to more relevance to country needs and the quality of the interventions delivered. Where partners' programmes are demand-driven, they have been swift to respond to requests, agree support and delivered the support on time. Moreover, for each of the partners, there are efficiency factors that deliver value for money: the funding for the World Bank leverages additional money and existing country relationships and dialogue; the PAL Network funding leverages past experience and existing country relationships; and for ACER UK the funding leverages its expertise on global standardisation.

181. The D4FL commitment to work through country structures (whether government or civil society) has implications for how fast and predictably resources are used. Country pressure can contribute to fast use, but other linked factors -- like country capacity, leadership and staff turnover, and political factors -- can cause delays versus plans. However, there is evidence that partners leverage positive factors, and take steps to mitigate negative factors. Moreover, internally, management of the programme at FCDO level and partners' level, supports efficient use of resources.

182. **Effectiveness:** The discussion above on relevance has already highlighted misalignment between the activities supported by the D4FL and the programme ToC. Some posted ToC outcomes and outputs are not well supported (inclusion of EiEPC, use of data, use of assessments in classrooms), and some activities are not well articulated in the ToC (the Accelerator programme), even where they contribute to the ultimate goal of improved foundational learning outcomes. There is mixed evidence on the extent to which D4FL funded activities adhere to the FCDO business case guiding principles for *Generate* and *Use* activities. There is more evidence to support principles of government ownership and building state capability, than on including teachers and school leaders and providing cost-effective sustainable solutions.

183. There are no warning signals with regards to programme implementation being stalled, so that the D4FL cannot be effective. Partners are making expected progress on delivering activities and

outputs. There are early promising results on more countries reporting globally, and the possibility that the programme will demonstrate that, at least for the countries where it is active, globally comparative data can be collected while also delivering data for national use. The issue is whether the net selection of countries across partners, and what individual partners do, provide sufficient proof for the programme to deliver a convincing result in this regard, by 2027.

184. In conclusion, the programme faces challenges to deliver its objectives. These are about design and the extent to which the programme setup facilitates achieving all its objectives, rather than about whether the individual partners are likely to deliver on their commitments and achieve the results they are aiming for. Even so, the programme is likely to increase the number of countries reporting, it is likely to improve data generation for national purposes in several countries, and can contribute to better learning in the countries where the Accelerator supports reforms.

185. **Looking ahead at impact and sustainability:** The EQ on impact and sustainability are about guiding the design of the summative evaluation. Even so, the findings provide lessons for the D4FL on programme design and implementation. Firstly, systematic achievement of ToC outcomes and impact is unlikely within the timeframe of the programme, because country assessment and reform cycles take time. This means that for the summative evaluation, the team will need to focus on early indicators, such as concrete country actions toward institutionalised demand for learning data; countries identifying assessment needs and developing country protocols for conducting a suite of quality learning assessments; different levels of government progressing toward the use of data to use in processes and reforms to improve foundational learning, and the links between these.

186. The evaluation found that partners' plans included some conceptualisation and integration of avenues to sustainable results, but not consistently across partners. Positive aspects are the focus on capacity strengthening, emerging efforts to support dissemination of results to stimulate demand for learning improvement, awareness raising activities at community level, and seeking government buy-in at different levels, including from teachers, to create the best possible conditions for sustainable impact. At the same time, sustainability can be integrated better in the programme design, for example through encouraging more focus on institutionalising demand for learning data and diagnostic capacity, and more efforts to reduce fragmentation on learning assessments.

187. The evaluation also found that the FCDO's role in the global assessment landscape is important, and could be leveraged by continuing to promote collaboration and disseminating the lessons from D4FL.

Summary conclusion

188. Across these criteria, and in answering the TOR overarching question, the evaluation therefore concludes:

189. **Some pathways are well designed individually, and early indication is that they are implemented well.** The UIS/GEM-R *Track* pathway delivers the global standards and public information goods. The ACER UK/ALMA and PAL Network pathways will each contribute to filling global learning data gaps incrementally, even with some inclusion of data from emergency contexts.

190. **Other pathways have significant gaps:** the pathway to policies, targets and interventions using learning assessment to strengthen curriculum, assessment and instruction has few activities supporting it, and not necessarily in a strategic way. There are also some gaps in the *Track* and *Generate* pathways, particularly on how systematically the programme is contributing to assessment in EiEPC contexts, to improve learning data for affected populations.

191. **There is, however, key design and implementation shortfalls on how well the programme is working together towards programme objectives.** A significant issue for the programme, however, is how well the different strands are balanced and are working together, and whether the use of resources is optimal, to deliver programme results. The evaluation concludes three main things: (i) Programme investment is skewed towards *Track*, while *Generate* and *Use* is more relevant for impact. (ii) The integration of *Track* with *Generate* and *Use* is not well articulated to serve the programme objectives. Key questions are how will *Track* contribute to *Generate* and *Use*, and what is needed to make the high investment in *Track* count for *Generate* and *Use*. (iii) There is little scope for the FCDO to influence countries selection and interventions and not enough coordination between partners to ensure programming choices are strategic for the D4FL.

5.2 Recommendations

192. The recommendation section answers the second overarching question for the evaluation: *What changes should FCDO and Implementing Partners make to increase the likelihood of programme success?*

193. We have six main recommendations. The first three relate to strengthening programme design. Of these, the first two deal with adjustments to the programme's ToC to articulate better how the programme will work together. The last is about trimming the programme down, so that it can focus on the things it already is funding and leverage them better. The final three recommendations are about how the FCDO and partners implement the programme.

194. We present the recommendations here in summary. An explicated table is provided in Annex 10, with more detail on the steps, the rationale, implementation responsibilities and timing.

Design recommendations

Recommendation 1. Adjust the ToC to reflect current programme design and intended mechanisms of change

195. **Rationale:** An accurate ToC will enable the summative evaluation to accurately evaluate the programme against its intended outputs and outcomes. In addition, the ToC can serve as a programming tool, helping partners understand how the different strands of work fit together, and enabling learning conversations about how well the different strands are progressing and how they might better join up.

196. **Proposed action:** Adjust the Theory of Change to reflect all the currently intended D4FL inputs, outputs, intermediate outcomes, outcomes and impact, and how they are expected to lead to one another. Annex 4 Figure 2 provides an adjusted D4FL ToC as a base for discussion.

197. **Responsibility and time frame:** The FCDO, for immediate implementation.

Recommendation 2. Articulate the relative importance of national and global programme objectives

198. **Rationale:** At the moment, the filling of global learning data gaps and the availability of disaggregated learning data that can be used to strengthen education provision in country are given equal weight in the Theory of Change (see Annex 4). *Generate* activities are more oriented to *Track* than to *Use*. Needs at country level can be conflicting with global reporting requirements. There is also ambiguity whether "filling global learning gaps" means eligibility under UIS criteria, or a broader category that is internationally comparable. A clear prioritisation of goals will bring clarity on where additional activities are needed. It will also support better evaluation.

199. **Proposed Action:** Consider and decide/articulate more clearly the relative priority of countries having any useful data on foundational learning vs. meeting global reporting requirements.

200. **Responsibility and timeframe:** The FCDO, for immediate implementation

Recommendation 3. Remove classroom use of assessment as an intended activity strand and shift to a focus on teachers as a key stakeholder group for D4FL *Generate* and *Use* activities

201. **Rationale:** The current ToC includes “Strengthened use of simple foundational learning assessment in classroom and community interventions”. However, there are currently no planned activities to bring about this output. This would be a substantial new strand of work and would likely require bringing in new partners to ensure the right capabilities would be included. It is unlikely that the time and budget remaining for the D4FL would be make substantial progress towards results in this area.

202. **Proposed Action:** Reorientate thinking away from use of assessment in classroom and community interventions, towards prioritising teachers as a key stakeholder group in national assessment systems. This will have implications for D4FL programming, and requirements for/guidance on partners’ activities on the ground. We recommend that FCDO and relevant implementing partners have a workshop in early 2025 to consider how teachers can be made more central in project efforts and what will be feasible within the time and resources available for the remainder of the project.

Responsibility and timeframe: FCDO, immediate; Partners, 2025-2027 in designing and implementing activities

Implementation recommendations

Recommendation 4. Increase the focus on the inclusion of EiE settings as a core programme objective

203. **Rationale:** The programmatic ToC forefronts emergency contexts in intermediate outcomes and outcomes. But this was not identified as a particular focus for implementing partners. Moreover, the incidental D4FL programme activities and delivery models are likely to be suitable for some emergency contexts (where displaced populations are internal or the state has responsibility for them) but not others (e.g. education for these populations are provided by non-state actors).

204. **Proposed Action:** Implementing partners to increase intentional focus on EiE, including for example working with in-country stakeholders to consider how conflict-affected areas or camps might be included in data collection exercises and ensuring tools are appropriate for such contexts. FCDO and partners to consider opportunities for working on foundational learning measurement in emergency contexts, even where they might require different working models (e.g. not working through government). FCDO and partners to work together to develop a typology for assessing EiE contexts, to identify what will be feasible, which in-country stakeholders to work with, and what additional expertise might be needed. FCDO to link implementing partners to resources available on conducting assessment in emergencies and to work with global partners on improving the architecture and evidence available on learning data generation in emergency and crisis contexts.

205. **Responsibility and timeframe:** FCDO, immediate; Partners, 2025-2027

Recommendation 5. Improve joint thinking and cooperation vertically between FCDO and partners and horizontally between partners at the global and country levels

206. **Rationale:** A clear strategy by the FCDO on how it will influence programme direction, and closer cooperation between partners (including all Track, Generate and Use partners) will contribute to the programme's strands working together better. The recommendation is based on the conclusion, and supporting findings, that while each strand of the programme will contribute to programme results (impact and outcomes) individually, the strands are not working together well enough to optimize how the D4FL investments, will achieve the outcome results.

207. **Proposed Action:** The FCDO should think strategically about how it can influence all partners to pull together to make progress. FCDO and partners should develop shared principles on coordination and cooperation that can guide expectations and responsibilities going forward. The FCDO should facilitate more joint thinking and cooperation between partner at the global level, on how to achieve the programme objectives jointly, and address shared challenges. Partners would benefit from investing in cooperation.

208. **Responsibility and timeframe:** FCDO on strategy and principles, immediate; FCDO and partners on joint thinking and coordination, 2025-2027

Recommendation 6. Partners should actively strengthen how they work politically to shift political and institutional barriers to regular assessment and use of data, and the FCDO and partners should invest in joint learning on how best to do so.

209. **Rationale:** Ensuring assessments are sustainable requires addressing political and institutional barriers to countries regularly undertaking assessments and using the resulting data to design and implement reforms in the instruction core of foundational learning. This recommendation is aimed at leveraging existing D4FL partnerships and contracted activities, to strengthen the links between Track – where a lot of investment is -- and Generate and Use, the engine room of learning improvement.

210. **Proposed Action:** Partners should strengthen whether and how their interventions will address existing and potential institutional barriers to imbedding learning assessments and using data, by shifting how assessments and learning data are seen and used. The FCDO and partners should invest in learning on how partners have successfully worked politically in this way.

211. **Responsibility and timeframe:** FCDO and partners, 2025-2027

Annex 1 Terms of Reference

OFFICIAL

Foreign Commonwealth and Development Office

ITT Vol. 2

ITT_6396

Evaluation of Data for Foundational Learning (D4FL) Programme: Terms of Reference

OFFICIAL

Contents

Acronyms List..... 3

1. Introduction..... 4

2. Context..... 4

3. Objectives of the Contract 5

5. Recipient and Beneficiaries 5

4. Scope of the Contract..... 6

6. Evaluation Approach and Questions..... 7

7. Contract Requirements..... 9

8. Skills and Competency Requirements..... 12

9. Performance management..... 12

10. Reporting Requirements 13

11. Budget and Payment Mechanism..... 15

12. Scale Down 16

13. Data and Risk Management..... 16

14. Governance, Quality Assurance and Contract Management Arrangements..... 17

15. Equity..... 17

16. General Data Protection Regulation..... 18

Appendix 1 of Call-down Contract (Terms of Reference): Schedule of Processing, Personal Data and Data Subjects 19

Annexes: 20

Acronyms List

ACER	Australian Council for Education Research UK Ltd.
BESTA	Better Education Statistics and global Action to improve learning programme
Coalition	Global Coalition on Foundational Learning
EdGE	Education, Gender and Equality Directorate
EQUALS	Evaluation Quality Assurance and Learning Service
FCDO	UK's Foreign, Commonwealth and Development Office
GAML	Global Alliance for Monitoring Learning
GED	Girls' Education Department
GEM-R	Global Education Monitoring Report
LICs	Low-income countries
LMICs	Lower-middle income countries
PAL Network	People's Action for Learning Network
SDG4	Sustainable Development Goal 4 – achieving 'equitable and quality education for all by 2030
SRO	Senior Responsible Owner
TA	Technical Assistance
ToC	Theory of Change
ToR	Terms of Reference
UIS	UNESCO Institute for Statistics

1. Introduction

1. The Foreign Commonwealth and Development Offices (FCDO) Girls' Education Department (GED) sits within the Education, Gender and Equality Directorate (EdGE) which leads on the delivery of the Government's commitment to stand up for the right of every girl around the world to 12 years of quality education.
2. This Terms of Reference (ToRs) details the requirements and responsibilities of a Supplier to provide an Evaluation of GED's centrally managed programme Data for Foundational Learning (D4FL).

2. Context

3. The D4FL programme is a new centrally managed programme that will provide up to £27.4m over four years to increase the amount, quality, comparability and use of learning assessments in primary schools in low- and lower-middle income countries (LICs and LMICs). The programme is solely funded by FCDO and is a successor to the Better Education Statistics and global Action to improve learning (BESTA) programme.
4. The D4FL programme works across LICs and LMICs and will have specific Technical Assistance initiatives in approximately seven countries and will be active in around 10 other countries through the work of the People's Action for Learning (PAL) Network. These countries are in Sub-Saharan Africa and South Asia. The programme is also active globally through its "track" component.
5. D4FL is central to the UK's commitment to women and girls as set out in the International Development Strategy and the recently announced Women and Girls Strategy.¹ It will directly contribute to Sustainable Development Goal 4 (to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all) by measuring and using learning data to improve education provision.
6. Most children in LICs and LMICs are in school, but they are not learning. On current trends, Sustainable Development Goal 4 (SDG4) – achieving 'equitable and quality education for all by 2030' – is unlikely to be met. To address this G7 leaders adopted two new, ambitious SDG4 milestone targets under the 2021 UK G7 Presidency:
 - 20 million more girls reading by age ten (or the end of primary school); and
 - 40 million more girls in school, by 2026.
7. Currently, there are significant gaps in the disaggregated and comparable data on learning needed to track progress against the G7 targets and SDG4.

¹ [International Development Strategy; Women and Girls Strategy](#)

More importantly, in poorer countries, reliable information on what children are learning, especially in primary school, is not collected frequently or at all, limiting the 'feedback loop' to national education policy and planning, and within the classroom itself.

8. D4FL has three components: track; generate; and use.
 - **Track** children's learning outcomes at global level to enable reporting on the SDG4 indicators for basic literacy and numeracy and the G7 global milestone targets for girls' education. This component is delivered through ongoing support to UNESCO Institute for Statistics (UIS) and the Global Education Monitoring Report (GEM-R).
 - **Generate** more learning data and improve the way national governments report children's learning outcomes through supporting the adaptation and implementation of the best available learning assessment tools in a range of partner countries. Support will be offered through technical assistance (current agreements with the Australian Council for Education Research UK Ltd. (ACER) and the World Bank) and promoting citizen-led learning assessments through the People's Action for Learning (PAL) Network.
 - Ensuring more learning data is **used** by policy makers in government and partners to underpin the design and implementation of policy, plans and interventions aimed to improve foundational learning in poorer countries (initially through the World Bank's work).

3. Objectives of the Contract

9. The purpose of this Contract is to provide an evaluation of the D4FL programme that will provide both learning and accountability. Overall, the objective is to assess the extent to which the design and implementation of D4FL supports its Theory of Change (Annex 2), specifically the two Outcomes:
 - More LICs and MICs report internationally using comparable and disaggregated data on learning, including in emergency contexts.
 - More state capability to regularly use disaggregated learning data so more governments and teachers implement and monitor interventions to improve foundational learning for all in priority countries.

5. Recipient and Beneficiaries

1. The recipient of this evaluation is FCDO.
2. The beneficiaries Implementing Partners, and other stakeholders in the global learning assessment architecture, such as members of the Global Alliance for Monitoring Learning (GAML)² and colleagues in Pillar 2 of the Global Coalition

² <https://gaml.uis.unesco.org/about/>

for Foundational Learning (the Coalition).³ Evaluation deliverables will also be of interest to and used by governments and development partners with whom the Supplier will interact to collect data and share learning.

3. Contract Duration and Value

10. The Contract will run for up to 34 months between June 2024 and March 2027. The Contract Financial Limit is up to a maximum of £490,000 inclusive of all local government taxes but exclusive of UK VAT.

4. Scope of the Contract

5. The Supplier will conduct formative and summative evaluation activities:

- Under the formative evaluation, the Supplier will evaluate the extent to which the structure, organisation and delivery of the programme is working as intended, making recommendations for how it may be adapted to perform better.
- Under the summative evaluation the Supplier will be required to evaluate the whole programme, assessing the extent to which it is contributing toward an increase in the amount of learning data available and its use to drive learning at scale.

6. The Supplier will be working with the Implementing Partners delivering the D4FL programme and with programme beneficiaries, which comprise policy makers and governments working in the education sector in supported LICs and LMICs.
7. Most evaluation activities will be desk-based, with some travel required to visit countries from the list of target countries, all in South Asia and Sub-Saharan Africa, to collect data and inform evaluations. These visits will principally inform the summative evaluation and focus on projects within the 'generate' component. FCDO requires a maximum of 10 country visits by the Supplier, the exact number and location of these visits will be determined once the Contract commences, during the inception phase.
8. The Supplier shall be responsible for the formative and summative evaluations covering all aspects of the D4FL programme but exclude evaluation of wider work by Implementing Partners that is not directly connected to the delivery of the D4FL programme. The D4FL evaluation will draw on the recently conducted independent evaluation of GEM-R⁴ and will focus on the interaction between this component ("track") and the rest of the programme. Evaluation of work by Implementing Partners under the other two components ("generate" and "use") will need to be more thoroughly evaluated since published evaluations are not available.

³ The Coalition for Foundational Learning was founded in 2022 to bring together partners with a shared commitment to improving foundational learning for all and a desire to work together to drive change more quickly. Members are Bill and Melinda Gates Foundation, FCDO, Global Partnership for Education (GPE), UNICEF, UNESCO, USAID and World Bank.

⁴ [Evaluations | Global Education Monitoring Report \(unesco.org\)](https://unesco.org)

9. The formative evaluation will be published and used by the FCDO D4FL programme team to triangulate progress reports from partners, and by FCDO and Implementing Partners to assess progress and adapt the delivery of the programme, if required. The formative evaluation will be completed before the second Annual Review of the programme in February 2025.
10. The summative evaluation will be published and used by FCDO to help fulfil its accountability requirements and for learning by FCDO and other stakeholders, helping guide decisions around investment in learning and informing future FCDO programming through internal and external knowledge sharing sessions.
11. The Supplier will undertake a mapping exercise during the inception phase to ensure that the appropriate individuals within each stakeholder category are included in dissemination efforts. The Supplier is required to include a full dissemination and use plan in the Inception Report.

6. Evaluation Approach and Questions

12. FCDO requires the formative and summative evaluations to take a theory-based approach using multiple methods suited to the evaluation objectives and questions. The formative evaluation will be a process evaluation using standard data collection methods such as document review, interviews and surveys. Suitable approaches for the summative evaluation might include contribution analysis, utilisation-focused evaluation and outcome harvesting. Use of comparison or control groups is not required.
13. FCDO is not prescribing a methodology for the conduct of the evaluations and these are only suggestions.
14. The formative evaluation will include all activities and projects in implementation whilst the summative evaluation will blend breadth and depth. It will encompass both a portfolio approach including all activities and projects and a deep dive into a selection of projects.
15. Evaluation questions for the formative and summative evaluations are set out below. The Supplier will work with FCDO during the Contract Inception Phase to finalise the evaluation questions and sub-questions.
16. The key evaluation questions for the formative evaluations are:
 - Is each aspect of the programme being implemented well, and working together to lead to the intended programme outcomes?
 - Could FCDO and Implementing Partners make changes to increase the likelihood of programme success?
17. These questions can be broken down into the following sub-evaluation questions aligned with the OECD-DAC evaluation criteria:⁵

⁵ <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

- a. **Relevance:** To what extent are D4FL's activities and objectives responding to the needs of country and partner institutions? To what extent are beneficiaries and key stakeholders engaging with the D4FL programme? Is the programme working to improve equity, through (1) supporting better understanding of gender differentials in education outcomes, and (2) ensuring more data is available on Children with Disabilities, and in Education in Emergency and Protracted Crises contexts?
- b. **Coherence:** To what extent does the structure of the D4FL programme, including relations between Implementing Partners, governance arrangements, and FCDO headquarters and country posts, support the delivery of the programme and achievement of its objectives? How does the programme support international learning assessment efforts, including work of key groups such as GAML and the Coalition? Are in-country programme objectives and activities aligned with national education structures and policies?
- c. **Effectiveness:** How successfully is each D4FL Implementing Partner working and are they making the expected progress in delivering their activities and outputs? What might be changed to increase the likelihood that the programme will be effective, including in its equity objectives? What is causing any observed variation in delivery of technical assistance and capacity building and are there more or less successful models of implementation?
- d. **Efficiency:** To what extent is the D4FL programme making best use of its resources to maximise the chance that programme outcomes will be achieved? To what extent are the resources allocated to programme activities appropriate and have these resources been used as planned? Is the programme being delivered according to the intended timeline?

18. The key questions for the summative evaluation are:

- Did the D4FL programme produce the intended results?
- Did it represent good value for money?

19. These questions can be broken down into the following sub-evaluation questions aligned with the OECD-DAC evaluation criteria:

- a. **Relevance:** To what extent did the activities and outputs funded through D4FL respond to the needs of country and partner institutions and other key stakeholders? To what extent did stakeholders engage with and support the D4FL programme? Did the programme support international learning assessment efforts, including the work of key groups such as GAML and the Coalition?
- b. **Effectiveness:** To what extent has investment through D4FL led to increased amount, quality, comparability and use of learning assessments in primary schools in D4FL target countries? Has the number of countries reporting against SDG4.1.1a and / or SDG4.1.1b increased? Has D4FL improved understanding of challenges and opportunities for girls' education? Has the programme led to an improvement in the availability and quality of learning data amongst

Children with Disabilities, and in Education in Emergency and Protracted Crises contexts?

- c. **Impact:** Has there been any change to government policies or practices with respect to foundational learning based on data made available through the programme? Has there been an increase in the number of countries with future plans to carry out regular robust assessments?
- d. **Sustainability:** What has the D4FL programme changed in the international assessment landscape? Will these changes continue after the programme has finished? How has practice changed in governments of countries who benefitted from the programme and are these changes likely to be sustained?

20. The Supplier will work with FCDO during the Contract Inception Phase to finalise the evaluation questions and sub-questions.

7. Contract Requirements

21. The Supplier shall deliver a set of outputs/deliverables and associated activities in line with the timings set out below.

22. The Contract will consist of two phases: an Inception Phase and an Implementation Phase.

Inception phase (10 weeks):

During the Inception Phase the Supplier will deliver the following:

23. The Inception Report is an Inception Phase deliverable and will need to be approved by the FCDO D4FL Senior Responsible Owner (SRO) before the Contract moves to its Implementation Phase.

During the Inception Phase the Supplier's deliverables will include, developing a workplan for the duration of the Contract, developing formative and summative evaluation plans, establishing a risk matrix and approach to risk mitigation, and writing a communication plan. The reporting requirements section below lists the contents of the Inception Report, which acts as a guide to activities that will need to be completed in this phase.

24. Engagement with the other D4FL Implementing Partners will be required during the Inception Phase to agree on data needs and timelines for evaluation activities.

25. A draft evaluation plan for the summative evaluation is to be included in the inception report, whilst the detailed planning for the summative evaluation to be further developed and be adapted during the Implementation Phase.

26. The final set of implementation deliverables, activities and timetable will be agreed with FCDO in the Inception Phase.

Output	Activities	Timeline
1. Draft Inception report	<p>Workplan for duration of the contract</p> <p>Formative and summative evaluation plans</p> <p>Risk matrix, including consideration of safeguarding.</p> <p>Communications plan</p>	Draft Inception report to be delivered 6 weeks after the contract start date.
2. Final Inception report	Based on activities in the draft Inception report	Finalised Inception report to be delivered 10 weeks after contract start date.

Implementation phase (33 months):

During the Implementation Phase the Supplier will deliver:

3. Formative Evaluation, which will provide evidence to FCDO on the degree to which implementation of the D4FL programme is working as intended and making recommendations for how programme delivery may be improved. The deliverables will be:
 - Presentation of draft findings to the D4FL Steering Committee. To be submitted in November 2024.
 - Organisation and facilitation of a learning and reflection workshop that draws on the formative evaluation and facilitates the adaptive management of the programme. This workshop will be in December 2024 and will include D4FL Implementing Partners.
 - A report against the evaluation questions set for the formative evaluation. The final report will be required in January 2025 to inform the second programme Annual Review.
4. Summative Evaluation, which will explore the extent to which the programme was successful in increasing the amount, quality, comparability and use of learning data in primary schools in LICs and LMICs. The deliverables will be:
 - A full draft summative evaluation plan submitted in September 2024 for review by the D4FL Steering Committee, with the final evaluation plan to be submitted by January 2025.

- Mid-term evaluation report summarising findings against the summative evaluation questions and making recommendations. This will be submitted in January 2026. It will be a learning-focussed report describing progress made towards the achievement of programme outcomes and making recommendations.
 - Final summative evaluation report submitted in December 2026.
 - Presentation of findings to relevant stakeholders (from the list of Beneficiaries, to be agreed between FCDO and Supplier) in February 2027.
5. With respect to the formative and summative evaluation outputs, presentations to the D4FL Steering Committee will be required. For details on this committee, see Governance and Quality Assurance Arrangements, below. The Supplier will also be required to produce a short summary of the summative and formative evaluation reports which FCDO will publish on their website.
6. The table below outlines the Implementation Phase deliverables and the associated activities along with the timelines:

Deliverable	Activities	Timeline
The Formative Evaluation	<ol style="list-style-type: none"> 1. A Field report showing evaluation with the chosen methods. 2. Final report on formative evaluation questions. 3. 2 workshops with (1) D4FL Steering Committee and (2) Implementing Partners, to support adaptive management. 	<p>Learning and reflection workshops in November and December 2024.</p> <p>Final written report in January 2025, to inform the D4FL Annual Review in February.</p>
The Summative Evaluation	<ol style="list-style-type: none"> 4. Final summative evaluation plan including evaluation framework, design, method, data requirements etc. 5. Conduct evaluation with the chosen methods. 6. Mid-term evaluation report summarising findings against the summative evaluation questions 	<p>Full draft summative evaluation plan submitted for FCDO review in September 2024.</p> <p>Final summative evaluation plan submitted January 2025.</p> <p>Mid-term report submitted January 2026.</p> <p>Final report in December 2026.</p>

	<p>7. Final report on summative evaluation questions.</p> <p>8. Presentation of findings to (1) D4FL Steering Committee and (2) with relevant beneficiaries (the Supplier may choose to run this as separate presentations for different partners)</p>	
--	--	--

8. Skills and Competency Requirements

- 7. The proposed team will need to possess the following skills and competencies to deliver the requirements of the Contract:
- 8. The Team Lead must have
 - a proven track record of leading evaluations in international development, in the field of education
 - expertise and competency in qualitative formative and summative evaluations and adaptive programme management
 - Expertise working with learning data and education in LICs and LMICs
 - Expertise in gender and inclusion of children with disabilities in education settings.
 - Ability to work collaboratively with a wide range of stakeholders working with LICs and LMICS, including FCDO officers, experts in education, international development programme implementers, and policy makers.
 - Excellent written and oral communication skills in English.
- 9. The team should also represent a mix of international and national/regional experts with considerations to a gender balance within the proposed team.
- 10. Local experts must be used for country-based field work (data collection activities) For field visits, participation of local evaluators is expected.

9. Performance management

- 11. FCDO will monitor the performance of the Supplier continuously for the entire duration of the Contract. FCDO will manage the Supplier’s performance through a range of following:
- 12. Regular (virtual) meetings with the Supplier every two weeks during the Inception Phase and then quarterly thereafter. Quarterly progress reports and

an annual report will be required to measure the Supplier against progress towards deliverables, and to ensure an agreed approach. An outline of the contents of these reports is set out below.

13. The Contract will have a dedicated FCDO Contract Manager and a Senior Responsible Owner (SRO) in FCDO, who will be the primary contacts for agreeing workplans and who will be responsible for the day-to-day management of the Contract. In addition, the Supplier will be in contact with other FCDO staff during the Contract.
14. All reports are to be submitted to the FCDO Contract Manager.
15. Reporting requirements, which may be adjusted in the Inception Phase, include:

10. Reporting Requirements

Inception Report

The Inception Report must be submitted ten weeks following the Contract commencement date. A draft report is to be submitted no later than 6 weeks following the Contract commencement date. (Arial font size 12, no longer than 30 A4 pages)

Inception Report must contain:

- A summary of progress and achievement against the Inception Phase deliverables.
- A refined/finalised workplan for the first year and an indicative workplan for the remainder of the contract, including a timetable, budget forecasts and a financial management plan.
- An evaluation plan and evaluation matrix detailing the questions and the approach for the formative evaluation.
- A draft evaluation plan and evaluation matrix for the summative evaluation.
- Format and timetable for the learning and the reflection workshop.
- Methodological limitations.
- Approach to equity.
- An outline of governance and quality assurance processes.
- A risk matrix including delivery chain risk map and contextual and methodological challenges for implementing the evaluations, with mitigation measures.
- A stakeholder map and stakeholder engagement / influencing strategy.
- Data sharing plans and agreements.
- An approach to ethics and safeguarding.
- An audience analysis and dissemination, use and influence plan for evaluation deliverables.
- A communications strategy outlining how the Supplier will represent themselves to FCDO and work alongside other D4FL Implementing Partners.

- A logframe covering the Supplier's work under this Contract, capturing the Supplier's own performance and milestones.

Quarterly Report

A Quarterly Report shall be due throughout the implementation phase. The fourth quarterly report will be replaced by the Annual Report. No longer than 10 A4 pages.

The Quarterly Report must contain:

- A narrative report on progress across the preceding quarter, including progress against the workplan and, when appropriate, results framework indicators.
- Planned activities for the coming quarter.
- Risk management update and updated Risk Matrix.
- Lessons Learned.
- Financial report including:
 - Reconciliation of previous quarter's expenditure against forecast.
 - Justification for variances of actual spend against the forecast.
 - Update of forecast of the remainder of the year.
 - Draft Quarterly Invoice for approval prior to submitting for processing.

Annual Report

An annual report must be submitted in January each year. This is timed to feed into the D4FL Annual Review. No longer than 20 A4 pages.

The Annual Report must contain:

- Summary of progress against the workplan and logframe for the previous year.
- Lessons learned and adaptations made to the evaluation approach, if appropriate.
- Progress against D4FL Annual Review recommendations relevant to the Suppliers work.
- Work plan for following year.
- Update on stakeholder engagement/influencing.
- Updated Risk Matrix including Supply Chain and mitigations.
- Financial update, setting out actual expenditure to date and revised forecast for the following years.
- Updated asset register

Formative Evaluation Report

A draft formative evaluation report must be submitted in November 2024. The final formative evaluation report must be submitted in January 2025. No longer than 40 pages excluding annexes. In addition, a 2-page summary ("evaluation digest") of the final formative evaluation report will be required using the template supplied by FCDO.

- The Formative Evaluation Reports will cover the relevant evaluation objectives and questions, the methodology, findings, conclusions, and recommendations, and will have an executive summary. The exact content and format of these reports will be established and agreed with Supplier in the Inception Phase.

Summative Evaluation Report

The final Summative Evaluation Plan must be submitted in January 2025. No longer than 20 A4 pages (excluding annexes). A mid-term Summative Evaluation Report must be submitted no later than January 2026 and the final Summative Evaluation Report submitted in December 2026. Each report shall be no longer than 50 pages excluding annexes. In addition, a 2-page summary (“evaluation digest”) of the final Summative Evaluation Report will be required using the template supplied by FCDO.

The summative evaluation plan will include the evaluation framework, questions, design, method, data requirements and workplan.

The mid-term and final reports will cover the relevant evaluation objectives and questions, the methodology, findings, conclusions, and recommendations, and will have an executive summary. The exact content and format of these reports will be discussed with the Evaluation Panel but must be agreed with FCDO.

16. In line with FCDO's evaluation policy, all evaluation reports will be published, together with a management response setting out how the FCDO will respond to the recommendations.
17. All the reports will be discussed at D4FL Steering Committee meetings and require a presentation from the Supplier to summarise the key learnings and deliverables.

11. Budget and Payment Mechanism

During the Inception Phase all payments will be output-based and linked to the successful delivery and acceptance of the Inception Phase deliverables. The amount due is set out Annex B Schedule of Payments
All expenses shall be paid on actuals.

18. Payment for the implementation phase will be made quarterly on the basis of input fees and actual expenses incurred in the period.
19. FCDO approval of the fulfilment of contract deliverables will assess the quality of each deliverable. All key contract deliverables will be reviewed by the FCDO SRO in accordance with this approach.
20. The supplier will submit payment invoices to FCDO quarterly in arrears. Invoices will illustrate fees and expenses incurred against the programme deliverables.

12. Scale Down

21. FCDO reserve the right to scale down or terminate this Contract at any point in line with the Terms and Conditions. Scaling down is at FCDO's discretion and may occur for reasons including but not limited to Supplier performance or shortage of funds.

Review Points

22. The Contract will be subject to a break point at the end of March 2025. Continuation of the Contract beyond that point will be subject to satisfactory performance and progress against the agreed deliverables. This will involve a review 6 months before the break point to determine whether performance and achievements to date are sufficient to continue and the Supplier will be notified no later than 2 months before the break point whether the Contract will continue into the next Spending Review period or be terminated at the break point.
23. The Contract will have a formal review point at the end of the Inception Phase to determine whether the Contract should proceed to the Implementation Phase. This will take place within two weeks of receiving the final inception report. Continuation of the Contract beyond the review point will be subject to acceptance of Inception Phase deliverables by FCDO.

13. Data and Risk Management

24. Most of the data required for this evaluation will be primary data generated by Supplier. Some additional data will come from publicly available sources such as the SDG4.1.1a and SDG4.1.1b UIS databases.⁶ The UIS is the official source of internationally comparable data on education. The evaluation will also use monitoring and evaluation data from Implementing Partners. This includes external evaluation data for GEM-R.⁷
25. When the Contract is signed, FCDO will set up an introductory meeting(s) with the Implementing Partners during the Inception Phase. This meeting will set out the evaluation design and establish data sharing principles and expectations, including consideration of timeliness of data sharing and collection throughout the programme cycle.
26. Successful engagement and data collection from representatives of national governments may require assistance from D4FL Implementing Partners and the FCDO D4FL Programme Team and local FCDO officers, for example to initiate contact and arrange interviews.

⁶ sdg4-data.uis.unesco.org.

⁷ [Evaluations | Global Education Monitoring Report \(unesco.org\)](#).

27. A full risk assessment will be conducted by the Supplier during the Inception Phase. Some key risks identified by FCDO the Supplier must consider in their approach to deliver this contract shall include but not limited to:
- Building relationships at the right level across D4FL Implementing Partners and stakeholders to collect data and evidence for evaluation.
 - Ensuring personnel with key responsibilities for delivering the activities under this Contract have the required skills and expertise.
 - Determining the contribution of the programme to outcomes and impacts, compared to overall shifts in the availability, quality, and use of data globally.
 - Retaining continuity in the evaluation team, given the duration of the services and Contract.

14. Governance, Quality Assurance and Contract Management Arrangements

28. Evaluation work under this Contract will be overseen and managed by the D4FL Steering Committee. This comprises the D4FL Programme Team and the Girls' Education Department Evaluation Adviser. The overall coordinator will be the D4FL Programme Responsible Officer (SRO). The Steering Committee will provide first-line quality assurance of deliverables under this contract. The Supplier is not expected to attend Steering Committee meetings. The final formative evaluation report and final summative evaluation report will be quality assured by FCDO's Evaluation Quality Assurance and Learning Service (EQUALS).
29. FCDO will manage the Supplier's performance primarily through a logframe (or other results framework) and progress against an agreed workplan. Each year FCDO will produce an annual review of the Supplier's performance and progress as per FCDO standard rules. This review will be completed in February each year. All the reports stated/listed under the reporting section will be used as the basis for assessing Contract performance.
30. Communication with Implementing Partners is a crucial part of this contract. Therefore, once the contract is awarded an introductory meeting will be set up between FCDO, Implementing Partners and the Supplier to formally present the Supplier to Implementing Partners, explain the Contract and its objectives and clarify any points of concern. Contact between the Supplier and Implementing Partners should be carefully managed. The Supplier will copy relevant FCDO contacts on significant correspondence with Implementing Partners.

15. Equity

31. The focus of the programme is to collect learning data that will be used to improve education outcomes for all children. Learning assessments need to be representative and show the learning status of marginalised groups through data disaggregation, including gender and disability status. The

programme also tries to improve understanding of learning status of children out of school, and those in emergency and protracted crises contexts. The extent to which the programme is successful in achieving these objectives will be evaluated by the Supplier.

32. The Supplier must take an inclusive approach to the conduct of the evaluation through for example team composition and considering equity throughout the evaluation.

16. General Data Protection Regulation

33. Please refer to the details of the GDPR relationship status and personal data (where applicable) for this project as detailed in Appendix 1 (Schedule of Processing, Personal Data and Data Subjects).

Appendix 1 of Call-down Contract (Terms of Reference): Schedule of Processing, Personal Data and Data Subjects

This schedule must be completed by the Parties in collaboration with each-other before the processing of Personal Data under the Contract.

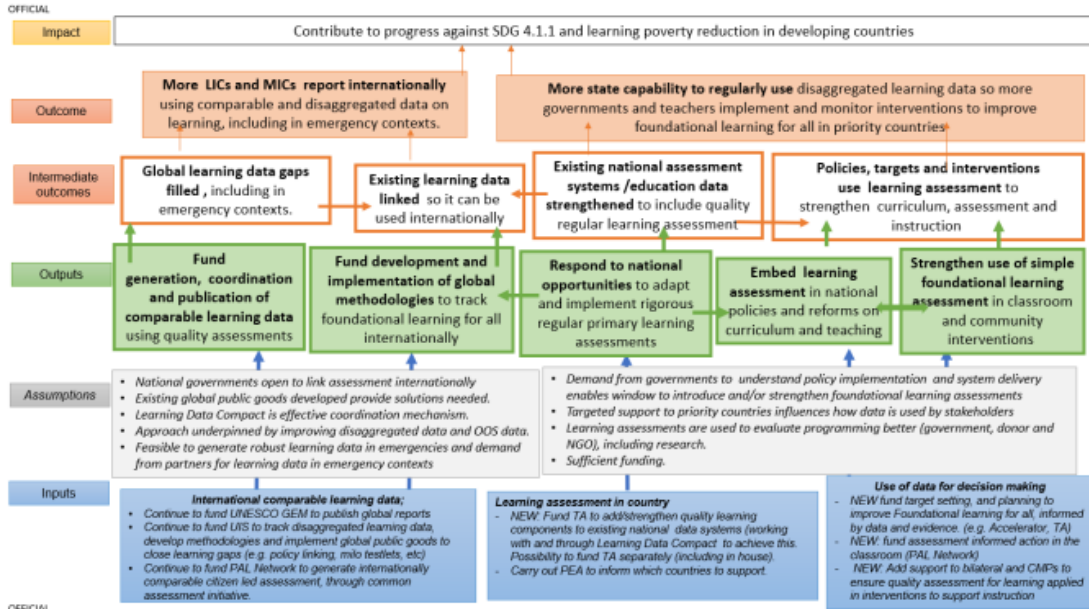
The completed schedule must be agreed formally as part of the contract with FCDO and any changes to the content of this schedule must be agreed formally with FCDO under a Contract Variation.

Description	Details
<p>Identity of the Controller and Processor for each Category of Data Subject</p>	<p>The Parties acknowledge that for the purposes of the Data Protection Legislation, the following status will apply to personal data under this Call-down Contract:</p> <p>1. FCDO is the Controller and the Supplier is the Processor in accordance with Clause 33 (Section 2 of the contract) of the following Personal Data:</p> <ul style="list-style-type: none"> • Names • Email addresses • Job titles

Annexes:

Annex 1: D4FL Logframe

Annex 2: D4FL Theory of Change



Annex 2 Formative Evaluation Methodology

Evaluation Audience and Stakeholders

1. The FCDO is the recipient of the evaluation and the principal user. The D4FL Programme Team was also principal informants for the formative evaluation. The FCDO education advisors and broader GED team are also principle users of the evaluation, and in the case of education advisors, important contact points and informants of the evaluation specifically in the country case studies.
2. Primary users and informants of the evaluation are the D4FL implementing partners, especially the World Bank, ACER UK and the PAL Network. The other partners, UIS and the GEM-R, were important informants and will also benefit from the evaluation, even if their programmes were evaluated only in respect of coherence with other *Track, Generate* and *Use* partners. The NFER, with which a contract was concluded late in 2024, will be a user of the evaluation, but an informant only for the summative evaluation.
3. Other stakeholders in the global learning assessment architecture, such as the FDO's peers in the Global Coalition for Foundational Learning, are also a primary audience, alongside the education authorities and other stakeholders, such as teachers and other assessment providers, in countries where the D4FL is active.
4. Secondary users of the evaluation are other national governments and other education stakeholders who may learn from the evaluation with reference to how best to support national and global actors to track, generate and use learning outcome assessment data. These actors are also of interest for the evaluation to collect perspectives from groups other than D4FL partners.

Evaluation questions

5. The overarching questions for the formative evaluation were:
 - Is each aspect of the programme being implemented well, and working together to lead to the intended programme outcomes?
 - Could FCDO and Implementing Partners make changes to increase the likelihood of programme success?
6. We have treated these as concluding questions, answering them through considering the 15 specific formative EQs set out in Table 8 below together. Table 1 in the main text maps the 15 EQs onto the evaluation criteria.
7. These questions were refined with the FCDO team during the inception period. The questions are a rationalisation of the questions in the TOR: where TOR questions are not reflected at the first question level summarised here, they are contained in the sub-questions and assessment criteria in the full evaluation framework, presented in Annex 3.
8. Due to the early stage of programme delivery and the formative nature of this evaluation, some questions are about what has happened to date while some are about what is *likely to happen* based on how the programme and support have been set up.
9. While relevance, coherence, efficiency and effectiveness are the most relevant evaluation criteria for the formative evaluation, EQs 13-15 look ahead to likely impact and sustainability, with the intention of informing the final evaluation framework for the summative evaluation.

Table 8 Formative Evaluation Questions

Evaluation criteria	Evaluation questions
Relevance	1. How relevant are the programme objectives to the needs of countries and the global community ¹³ ?
	2. How relevant is the programme’s design to achieving its objectives, including its equity objectives?
	3. To what extent are current processes likely to ensure countries’ needs are identified and met?
Coherence	4. To what extent is D4FL a coherent programme across strands and partners?
	5. To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies?
	6. To what extent are D4FL and other FCDO projects ¹⁴ complementing one another?
Efficiency	7. To what extent will the programme set-up likely facilitate timely and efficient use of resources?
	8. To what extent are partners responding in a timely way to requests for support?
	9. To what extent have D4FL resources been used efficiently to date?
Effectiveness	10. To what extent is the implementation of D4FL activities likely to lead to outcomes as set out in the programme ToC?
	11. What has the programme achieved to date?
	12. What are the likely barriers and facilitators to achieving results?
Looking ahead to likely Impact and Sustainability	13. What are the likely pathways and timelines for outcomes and desired impact?
	14. What are the avenues for D4FL to have sustainable impact?
	15. What does sustainability look like for D4FL?

Evaluation plan and matrix

10. Table 8 provides an overview of our methodology and sets out how our design intended to address the evaluation criteria and questions. Below we provide further detail on the nature and purpose of each data collection activity. In Annex 3 we provide a full evaluation framework for the formative evaluation, outlining sub-questions, assessment criteria and data sources for each EQ.

11. The formative evaluation included global- and country-level data collection. Below we provide further detail on how country cases were selected.

¹³ The global community refers to institutional stakeholders in the global learning assessment architecture. In the stakeholder analysis these are the FCDO, its D4FL partners (UIS, GEM-R, ACER UK, PAL and the World Bank), and institutions in the Global Learning Architecture such as UNICEF, GPE, BMGF, USAID, WB, UNESCO, ECW and INEE.

¹⁴ Projects in scope for now are the FCDO centrally managed education programmes, such as SCALE, the What Works Hub, EdTech, and bilateral programmes in the D4FL countries covered. The evaluation team may add further FCDO projects which emerge as relevant as it undertakes its work. The evaluation team will also look out for non-FCDO programmes that are relevant, particularly in the different country contexts for the *Generate* and *Use* streams.

Table 9 Evaluation matrix for the formative evaluation

	Global-level					Country-level			
	Document review	External assessment expert interviews	TOC workshops with partners	Partner interviews	FCDO D4FL lead interview	Document review	Government interviews	Partner interviews	FCDO adviser interviews
Relevance									
1. How relevant are the programme objectives to the needs of countries and global community?	X	X		X		X	X	X	X
2. How relevant is the programme's design to achieving its objectives, including its equity objectives?	X	X	X						
3. To what extent are current processes likely to ensure countries' needs are identified and met?	X		X	X	X	X	X	X	X
Coherence									
4. To what extent is D4FL a coherent programme across strands and partners?	X		X	X	X	X	X	X	X
5. To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies?						X	X	X	X
6. To what extent are D4FL and other FCDO projects complementing one another?	X		X	X	X	X	X	X	X
Efficiency									
7. To what extent will the programme set-up likely facilitate timely and efficient use of resources?	X			X	X			X	X
8. To what extent are partners responding in a timely way to requests for support?				X		X	X		X
9. To what extent have D4FL resources been used efficiently to date?	X			X	X	X		X	X
Effectiveness									
10. To what extent is the implementation of D4FL activities likely to lead to outcomes as set out in the ToC?			X	X		X	X	X	X
11. What has the programme achieved to date?	X			X		X	X	X	X
12. What are the likely barriers and facilitators to achieving results?	X	X	X	X	X	X	X	X	X
Looking ahead to Impact and Sustainability									
13. What are the pathways and likely timelines for outcomes and impact?	X	X	X				X	X	X
14. What are the avenues for FCDO to have sustainable impact?	X			X	X	X		X	X
15. What avenues for FCDO to influence the wider sector?	X	X		X	X			X	X

Sampling and data sources

12. We carried out a **global-level document review** in September 2024 to understand the context in which the D4FL is being delivered, assess plans for partners' delivery, identify external assessment experts to interview and countries to include in the country-level data collection (see below), and understand activities, results, and barriers, facilitators and risks to date. This review included the FCDO's and partners' plans and reports, documentation on which countries are currently reporting against SDG 4.1, and published literature on the global assessment landscape and previous assessment efforts.

13. We **interviewed 3 external assessment experts** who are not involved in the D4FL programme in November-December 2024, to understand the relevance of the programme to the global education community, and hear their lesson learnt on potential barriers, facilitators, risks and pathways to impact. These experts identified were individuals identified as having worked with governments to establish national assessment systems (e.g. Team Leads from initiatives listed in the UIS database) and/or who have established citizen-led assessments in LICs and MICs (e.g. RELI, ASER). The interviews aimed to capture their views on the needs and priorities of LICs, MICs and the global education community and lessons learnt from these past initiatives. We then assessed the extent to which the D4FL programme design and activities meet these needs and priorities and are likely to address previous mistakes and challenges.

14. We carried out a 2.5-hour, online **ToC workshop with each of the three partners** currently involved in delivering the *Generate* and *Use* strands (ACER UK, the PAL network and the World Bank) in October 2024.¹⁵ We engaged with our key point of contact for each partner to prepare an invitation list for the workshops, aiming to include strategic and operational team members. In the workshops, we facilitated each organisation to articulate a ToC for the support that they will provide and how it will lead to the ultimate intended outcome of contributing to progress against SDG 4.1.1 and learning poverty reduction in developing countries. This included articulating assumptions, including intended causal mechanisms, and how the support they provide will interact with other support and activities. This informed our explication of the D4FL ToC, and allowed us to consider how much coherence there is between the different programme strands and providers, and assess the extent to which providers' plans and thinking align with FCDO's programme theory and guiding principles.

15. We **interviewed global representatives of the currently contracted partners** across all three programme strands (ACER UK, the PAL network, UIS, GEM-R, and the World Bank) between October-December 2024, to capture their views on relevance, coherence and results achieved to date and to explore in more depth their plans and delivery to date.

16. We also **interviewed the FCDO D4FL lead** to capture reflections on the relevance, coherence, efficiency and effectiveness of the programme design and work carried out to date. In this interview we also looked ahead to Impact and Sustainability, considering potential and desired avenues to influence the wider sector, and exploring the ambitions for the programme's longer-term impacts in-country and globally. This interview took place in

¹⁵ For the World Bank, we carried out separate workshops for ALMA and the Accelerator. Each lasted 90 minutes.

November 2024 so that it could explore some emerging findings from the desk review and interactions with partners.

17. Following the initial document review and ToC workshops, the team identified **five out of the ten countries where work has started for country-level online data collection**, communicated a long list of countries to the FCDO in a country case study selection note. After discussion with the FCDO the team selected the final five countries, namely Ghana, Kenya, Nepal, Senegal and Somalia. In Ghana (ACER UK, World Bank ALMA and Accelerator), Nepal (World Bank ALMA, PAL Network), Senegal (PAL Network, World Bank Accelerator) and Somalia (ACER UK, World Bank ALMA), there are two D4FL partners operating, while in Kenya only the PAL Network is operating. The decision to select countries with more than one partner was made to explore coherence and synergies between partners' approaches.

Box 1 **Selecting the country cases**

The selection of country case studies for the formative evaluation of the D4FL Programme followed a purposive sampling strategy, balancing typical and outlier cases. Five countries were selected: Ghana, Kenya, Nepal, Senegal, and Somalia. These were chosen to reflect diverse contexts, partner involvement, and varying stages of progress in D4FL implementation.

Key considerations included:

- **Representation Across Partners:** Countries included contributions from all three D4FL partners (PAL, ACER UK, and World Bank), representing varying stages of progress and implementation. The overlaps in Ghana, Nepal, Senegal and Somalia provided opportunities to evaluate collaborative efforts and synergies between partners.
- **Contextual Diversity:** The sample captured a mix of contexts, including the protracted crisis in Somalia and Kenya which hosts a significant refugee population, variations in reporting on learning outcomes, and differing trends in education performance.
- **Regional and Linguistic Variation:** Countries spanned East and West Africa, as well as South Asia, and included English- and French-speaking contexts.
- **Potential for Summative Evaluation:** Ghana, Somalia, and Senegal offered strong potential for inclusion in the summative evaluation, due to overlapping interventions and contextual complexity.

18. We carried out the following activities for each sampled country, across October-December 2024:

- **Review of country-level plans and reports** to understand the support that has been agreed and progress made to date.

- **Interviews with Government or the key in-country stakeholders¹⁶** to understand their perspective on the relevance and value of the support provided to date and their experiences of the programme. These interviews were held online and lasted up to one hour. One interview per country was conducted during the formative evaluation, with the number of participants ranging from one to three, depending on the partnership structure in each country.
- **Interviews with the relevant partner organisation** working in-country. These interviews focused on relevance, coherence, efficiency and effectiveness within that specific country, including the factors that have shaped the support planned and provided and any anticipated opportunities and challenges going forward. The interviews were held online and lasted up to one hour. If the same individual within an organisation is overseeing work in multiple sampled countries, the conversations were scheduled flexibly to accommodate participants. Each interview typically involved one or two individuals who have led the relationship and the work in the sampled country.
- **Interview with the FCDO education adviser** for sampled countries where this was applicable. These interviews focused on the process for setting up the support and how F4DL activities are interacting with other activities in-country, as well as capturing broader reflections on relevance, effectiveness, efficiency and challenges and opportunities. These interviews took place online and lasted up to one hour.

19. Table 10 below summarises our respondent sampling approach and anticipated and achieved sample sizes for each primary data collection activity.

Table 10 Formative respondent selection for data collection

Activity	Intended sampling strategy	Intended n.	Achieved n.
Global level			
External assessment expert interviews	At least two each of individuals who have worked with governments to establish national assessment systems individuals or small teams who have established citizen-led assessments in LICs and MICs	6	3
Interviews with members of the GCFL	Not included in original strategy	0	1
ToC workshops with partners	One workshop per organisation (ACER UK, PAL, World Bank)	3	3
Partner interviews	1-2 interviews with strategic and operational team members per organisation (ACER UK, PAL, UIS, GEM-R, World Bank)	8	7
Learning and reflection workshops	1 Workshop with FCDO stakeholders, and 1 workshop with implementing partners	2	2
FCDO D4FL lead interview		1	1

¹⁶ For countries working with the PAL Network, the Network are likely to work with Civil Society Organisations (CSO) in the way that other partners might work with Government. For the remainder of this IR, references to Government participants refer to Government or CSO as applicable.

Activity	Intended sampling strategy	Intended n.	Achieved n.
Country-level (five sampled countries)			
Government interviews	One interview with Government representative(s) (or key in-country stakeholder(s) per sampled country. We did not secure an interview with government in Senegal and Nepal in time for the evaluation.	5	3
Partner interviews	One interview with ACER UK/PAL network/World Bank country lead(s) per sampled country. We did not secure an interview with World Bank education team in Ghana in time for the evaluation.	<9	8
FCDO adviser interviews	One interview per sampled country where FCDO adviser in post. We did interview advisor in Ghana who was on maternity leave.	<4	3

20. We recorded workshops and interviews with participants' permission, complemented by researchers taking notes. To encourage open and frank discussion with respondents, all respondents were assured that the interview recordings and notes would be available only to the evaluation team, that they would not be named or quoted without their written consent in the evaluation report, and that care would be taken in reporting so that respondents' identity cannot be inferred. For this reason there is not a list of consultees included with this report. In line with our commitments on ethics, safeguarding and data use and protection (see from paragraph 38 below), informants received a summary of the evaluation purpose, scope and approach before the interview alongside information on consent and data use. At the start of interviews, the team drew informants' attention to this, and reiterated how information shared will be used, that the data will be securely kept by the evaluation team until three months after the end of the evaluation, and that informants can indicate when they share information off record, or when information should be treated as highly sensitive.

21. The evaluation did not set out to produce country case reports. We summarised the data collected for each formative evaluation case in an analytical framework, as discussed in the next sub-section. This framework was internal to the team.

Analysis

22. We drew on the "Framework" approach^{to} manage the qualitative data (Ritchie, J. et al., 2013). We assembled a matrix in which each row represented an individual interview, group discussion or document and each column an evaluation sub-question or analytical theme. We summarised the qualitative data in the matrix, including illustrative quotes where appropriate. Once we had summarised all data in this matrix, we analysed it thematically. The Framework approach enabled us to carry out both descriptive and explanatory analysis, identifying convergence and dissonance between individuals and between participant groups and looking for explanations for differing views.

23. The quantitative data used in the formative evaluation was limited to secondary data from FCDO and partners, including financial reporting data and any quantitative monitoring indicators. This data was analysed descriptively, with measures such as averages, ranges, and proportions reported.
24. We carried out within- and across-case analysis. We first triangulated perspectives to carry out within-case analysis of each country. The aim of this within-case analysis will be to address the EQs for the individual countries, understanding their experiences and opinions and reasons before moving to cross-case analysis.
25. Our cross-case analysis then focused on identifying convergence and dissonance between countries and looking for explanations for what has happened and been said. This enabled us to identify patterns and draw conclusions for each EQ at a global level. It also enabled us to consider, for each question under Relevance, Coherence, Efficiency and Effectiveness, how findings differ between partners and delivery models, and what this might imply for the likely success of different models going forward.
26. Our analysis was continuous throughout the formative evaluation. For example, we analysed data from the partner ToC workshops so that we could follow up on emerging findings in later research activities. This analysis also informed our selection of countries for country-level data collection.
27. Throughout, we triangulated findings from different data sources to address the evaluation questions. We used an emerging findings matrix to summarise findings and evaluative judgements, synthesise the evidence from across sources that supports the finding, and assess the weight of that supporting evidence. This triangulation culminated in an internal analysis workshop, led by the Team Lead. The ET came together to analyse findings from different sources under each of the research questions, assess the weight of evidence and identify convergent and divergent findings, which we explain in our reporting. Where findings diverged between sources, we revisited the raw data and considered potential theoretical and methodological reasons for these differences. This form of peer discussion also served the purpose of quality assurance, ensuring legitimacy of findings (Education Endowment Foundation, 2022).
28. As part of our triangulation of findings, we “RAG rated” the assessment criteria set out in our evaluation framework (see Annex 9). We colour coded each criterion, with Red indicating that it did not happen or is at high risk; Amber that it only happened in some circumstances or for some elements of support or is at moderate risk; and Green indicating that it did happen or can be expected to happen with a high level of confidence.
29. For the summative evaluations, we will also use contribution analysis (linked to the relevant ToC) to determine the extent to which the D4FL programme, versus other actors, contributed to the achievement of the programme’s objectives, considering the country and global context for the programme. The contribution analysis will also use a matrix approach, to identify and make judgements on the strength of D4FL’s contribution to the results observed. We did not use this approach in the formative evaluation, as indicated in the IR, because of the early stage of programme implementation and because the country case studies were light touch.

30. Another output of our analysis of the formative evaluation data was an explicated, diagrammatic ToC for D4FL (see Annex 4). This will draw on the theories of change discussed in the ToC workshops with partners along with broader learning about how the programme is intended to work. It will include the assumptions and causal mechanisms underpinning the ToC for D4FL, including the links between programme strands and partners. It will be this explicated ToC that the summative evaluation goes on to test.

31. We shared emerging findings and the explicated ToC during the Learning and Reflection workshops with FCDO and partners before finalizing our analysis and moving to the reporting phase. The partners' workshop concentrated on five key findings from the formative evaluation, with a primary focus on identifying lessons the programme can learn from these findings and exploring potential changes to enhance its chances of success.

32. The outputs of the formative and summative evaluations are detailed in Table 11 below.

Table 11 Written outputs from the formative and summative evaluations

Output	Description	Timing (finalised)
Formative evaluation outputs		
Explicated ToC	A diagrammatic ToC with explicated assumptions and causal mechanisms, including links between strands and partners Figure 2 in Annex 4.	January 2025
Formative Evaluation Report	Including an introduction to the programme and evaluation, an outline of the evaluation questions and methodology, findings structured by evaluation criteria and evaluation questions, conclusions and recommendations. Conclusions reflecting the RAG rating of the assessment criteria set out in our evaluation framework, and including a narrative conclusion against each of the evaluation criteria (Relevance, Coherence, Efficiency, Effectiveness and looking ahead to Impact and Sustainability). Recommendations split into recommendations for the ongoing delivery of F4DL and recommendations for the summative evaluation design.	January 2025
Formative Evaluation Digest	A 2-page summary using the template supplied by FCDO and including key findings and implications for the ongoing delivery of the F4DL programme and for the summative evaluation design.	January 2025
Final Summative Evaluation Plan	Final evaluation plan, evaluation matrix and evaluation framework following the structures used in this IR. Any relevant updates to the IR, for instance around ethical considerations or data sharing plans based on the final methodology.	February 2025

Validation

33. The Learning and Reflection workshops provided a first layer of validation with the principal and primary stakeholders. In addition, the principal and primary stakeholders will be provided with an opportunity to undertake a fact check and raise queries on the draft formative evaluation report.

Approach to equity

34. We will mainstream equity considerations throughout the design, implementation and analysis of this evaluation.

35. In designing the formative EQs we considered the equity dimensions of the assessment and of the D4FL programme. This is explicit in some EQs – for example question 2 and sub-question 10.1 address the “equity objectives” of D4FL, which are that more LICs and MICs report internationally using data that is **disaggregated** and **includes emergency contexts** (*Generate* strand) and that there is more in-country capability to use **disaggregated** data to improve foundational learning **for all**. Moreover, also considered equity in our analysis of data for all questions – for example, which children are included and excluded from assessment data collection? What steps have been taken to ensure children from marginalised groups, including those with disabilities can be included? How is data intended to be disaggregated? And who has been included and excluded from the conversations about the purpose and nature of assessments that inform their design?

36. During in-person country fieldwork, we will also ensure representation of different marginalised groups when targeting our interviews and focus group discussions. This would mean, for example, ensuring that we talk to (I)NGOs active in the education sector and working on gender, children with disabilities, and internally displaced persons and refugees.

37. We have a named team member responsible for reviewing evaluation tools and outputs to ensure we attend to equity concerns across the evaluation. Below we outline further strategies to ensure that risks around gender insensitivity and power imbalances were mitigated in our interactions with participants.

Ethical considerations

38. Mokoro is responsible for safeguarding and ensuring ethical conduct at all stages of the evaluation cycle. This includes, but is not limited to, ensuring informed consent, protecting the privacy, confidentiality and anonymity of participants, ensuring cultural sensitivity, respecting the autonomy of participants, ensuring fair recruitment of participants (including women and socially excluded groups) and ensuring that the evaluation results do no harm to participants or their communities.

39. Mokoro’s approach to safeguarding the evaluation against ethical risks include:

- Providing Mokoro’s Code of Conduct to all evaluation personnel when contracted, and drawing their attention to the Code before signing the contract. The following Mokoro policies link to the Code of Conduct and are linked in all consultant contracts: Equal Opportunities and Diversity; Safeguarding; Whistleblowing; Anti-bullying and Harassment; and Anti-fraud, Bribery and Corruption. The Code includes issues of political and cultural sensitivity when engaging stakeholders. The team leader and country case study leaders will ensure that all team members are familiar with the Code of Conduct when work commences.
- Sharing an Evaluation Introduction Brief with key informants, providing information on the evaluation, what reports will be published and/or shared with them, how their data and information would be used and protected, and dealt with when the evaluation is

complete, and assurance that engagements would be confidential to the evaluation team.

- Deploying strategies to minimise the burden on stakeholders, namely carrying out a desk-based of documentation and data prior to interviews and workshops, so that primary data collection was focused on answering questions that cannot be answered through secondary data, keeping research interactions to a proportionate minimum length and offering flexibility in scheduling.
- Undertaking the evaluation with integrity, under oversight of the team leader with support from the deputy team leader. This includes ensuring that stakeholders are given feedback on findings, that findings are clearly and transparently accounted for, and that the report is balanced.
- Quality assurance internally and through the D4FL Steering Committee and the EQUALS process to ensure that evaluation outputs meet ethical and integrity standards.
- Working with the FCDO to ensure that dissemination of findings is adequate, appropriate and well balanced.

Team Roles and Responsibilities

40. There is no known conflict of interest concerning any team members. All core team members actively engaged in data collection and analysis activities, contributing valuable inputs to the evaluation outputs. The team leader held overall responsibility for managing evaluation processes in alignment with the IR, as well as overseeing the drafting, delivery, and finalisation of the formative evaluation report. The deputy team leader provided strong support to the team leader, taking the lead on selected processes and conducting the initial quality control of the country case studies.

41. The table below provides a more detailed overview of the team roles and responsibilities.

Team member and role	Responsibilities/coverage
<p>Alta Fölscher Team Leader</p>	<ul style="list-style-type: none"> • Team Leader with overall responsibility for all aspects of the evaluation • Primary responsibility for supervision of ET members at all phases of the process • Responsible for overall design, implementation, reporting and timely delivery of the evaluation process and products • Provided overall leadership and ensured quality process and products during the data collection phase • Led consolidation and analysis of the data collected during the evaluation, ensuring high-quality triangulated evidence was used in the response to all EQs • Led the team workshop to draw findings and conclusions together from the body of evidence collected during the evaluation • Oversaw team members' inputs to the ER
<p>Lydia Marshall</p>	<ul style="list-style-type: none"> • Overall support to the TL in ensuring the good conduct of the evaluation • Supported the TL in drafting of the ER

Team member and role	Responsibilities/coverage
Deputy Team Leader	<ul style="list-style-type: none"> Supported team leader on quality control of country case study outputs Co-led the team workshop to draw findings and conclusions together from the body of evidence collected during the evaluation Co-drafted the ER, and other products of the evaluation, ensuring quality standards and agreed timelines are met
Christine Fenning Evaluator	<ul style="list-style-type: none"> Provided evaluation and thematic expertise on education Participated in data collection phase Conducted relevant and selected interviews and other qualitative data collection Participated in the team workshop to draw findings and conclusions together from the body of evidence collected during the evaluation Reviewed the evaluation tools and outputs for effectiveness of mainstreaming equity (gender, children with disabilities and EiEPC affected population) concerns across the evaluation Inputted into the ER in line with their areas of responsibility
Additional Experts	
Principal expert Rachel Outhred	<ul style="list-style-type: none"> Provided technical expertise to the evaluation on measuring learning outcomes and improving educational effectiveness Participated in the team workshop to draw findings and conclusions together from the body of evidence collected during the evaluation Provides inputs into ER, ensuring quality standards and agreed timelines are met
Pool of technical experts Paulina Valenzuela	<ul style="list-style-type: none"> Provided technical expertise on learning assessment in education, and additional evaluation expertise and skills Participated in the team workshop to draw findings and conclusions together from the body of evidence collected during the evaluation
Research Support	
Tal Shalson Research Coordinator Sara Bashir Malik, Anushay Mazhar Additional research support	<ul style="list-style-type: none"> Focal point for liaison with FCDO Managed the team's e-library, and the (confidential) interview database Undertook literature review, data analysis, recording interviews, and assisted with scheduling of interviews Supported the team in coordination of interviews and focus group discussions. Participated in the team workshops to draw findings and conclusions together from the body of evidence collected during the evaluation Assisted in preparation and editing of draft and final reports, including ensuring EQUALS templates are followed by the team.
Quality Support	
Stephen Lister Quality Support	<ul style="list-style-type: none"> Ensured compliance with EQUALS system Advised on the evaluation findings, conclusions and recommendations Reviewed and commented on the main evaluation products before their submission

Team member and role	Responsibilities/coverage
Assignment support	
Céline Cornereau Assignment Manager	<ul style="list-style-type: none"> • Responsible for assignment administration, contracts and logistics; administrative liaison with client • Assessed and ensured effective management of risks, taking account of Mokoro's duty of care • Monitored assignment budget and invoicing • Coordinated administrative support to the assignment team • With the research coordinator, ensured that assignment deliverables meet the agreed design/presentational standards

Annex 3 Formative Evaluation Framework

FORMATIVE EVALUATION FRAMEWORK

FORMATIVE EVALUATION FRAMEWORK			
CONCLUDING / OVERARCHING QUESTIONS	Is each aspect of the programme being implemented well, and working together to lead to the intended programme outcomes? Could FCDO and Implementing Partners make changes to increase the likelihood of programme success?		
MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
CRITERION: RELEVANCE			
The extent to which D4FL's objectives, design and activities to date respond to and meet the needs of LIC/LMIC country-level and global policy makers on and stakeholders in education outcome monitoring and standardised learning assessments.			
1. How relevant are the programme objectives to the needs of countries and global community? - Objectives include the intermediate outcomes, outcomes and impact set out in the ToC	1.1 To what extent are the programme objectives relevant to the needs of LICs and MICs?	Country governments, teachers, community leaders and other country stakeholders are interested in and working with partners on activities in countries Countries lack learning data to inform policy and practice Countries understand the implications of and want to generate comparable and disaggregated learning data	Document review of partners' reports Document review of currently available evidence on learning outcomes in-country Global-level interviews with partners Country-level interviews with government, partners and FCDO advisers
	1.2 To what extent has support provided to date been relevant and aligned to countries' needs?	Countries have found the support to be relevant and aligned to their needs	Country-level interviews with government, partners and FCDO advisers
	1.3 To what extent are the programme objectives relevant to the needs of the global community?	There are global learning data gaps There is a global commitment to filling these gaps in order to monitor education quality and learning outcomes Global learning assessment stakeholders have engaged with the D4FL programme and are interested in aligning activities and/or making use of outputs	Document review to understand extent to which countries are currently reporting internationally against SDG 4.1 Interviews with external assessment experts Global-level interviews with partners and FCDO lead
2. How relevant is the programme's design to achieving its objectives, including	2.1 To what extent are planned activities designed to leverage opportunities and address barriers to	Planned activities are appropriate and contextually relevant to participating countries Planned activities address real opportunities and barriers to international reporting in LICs and MICs	Document review of partners' plans Document review of literature on global assessment landscape and previous efforts ToC workshops with partners

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
on equity?	international reporting of comparable and disaggregated learning data?		Interviews with external assessment experts
	2.2 To what extent are planned activities designed to leverage opportunities and address barriers to international reporting of learning data in emergency contexts?	Planned activities are appropriate and contextually relevant to emergency contexts Planned activities address real opportunities and barriers to international reporting in emergency contexts	Document review of partners' plans Document review of literature on global assessment landscape and previous efforts ToC workshops with partners Interviews with external assessment experts
	2.3 To what extent are planned activities designed to leverage opportunities and address barriers to raising the quality of national assessment systems and learning data?	Planned activities are appropriate and contextually relevant to participating countries Planned activities address real opportunities and barriers to all aspects of quality including utility, credibility, integrity, validity, reliability and fairness	Document review of partners' plans Document review of literature on global assessment landscape and previous efforts ToC workshops with partners Interviews with external assessment experts
	2.4 To what extent are planned activities designed to leverage opportunities and address barriers to governments and educators using learning data to improve foundational learning?	Planned activities are appropriate and contextually relevant to participating countries Planned activities address real opportunities and barriers to governments and educators using learning data to improve foundational learning, including for marginalised or disadvantaged groups, especially girls	Document review of partners' plans Document review of literature on global assessment landscape and previous efforts ToC workshops with partners Interviews with external assessment experts
3. Are current processes likely to ensure countries' needs are identified and met?	3.1 To what extent are processes for deciding who should support whom, and how, likely to ensure countries' needs are met?	FCDO and implementing partners have appropriate processes in place to decide who should support whom, and how, which will ensure countries are met To date, support is being provided by partners with the appropriate capacities to meet countries' needs	Document review of partners' plans and reports ToC workshops with partners Global-level interviews with partners and FCDO lead

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	3.2 To what extent are partners' processes adequately identifying and ensuring that TA meets countries' needs?	Implementing partners have appropriate processes in place to assess countries' needs Implementing partners are tailoring support provided/agreed to date to countries' needs, priorities and capacities	Document review of partners' plans and reports ToC workshops with partners Country-level interviews with government, partners and FCDO advisers
<p>CRITERION: COHERENCE</p> <p>The extent to which D4FL's objectives, design and activities to date are internally coherent and complementary, and align with country- and global-level education structures, policies and priorities and assessment initiatives.</p>			
4. To what extent is D4FL a coherent programme across strands and partners?	4.1 To what extent are D4FL partners working towards common objectives?	Implementing partners' theories of change align with the programmatic ToC FCDO, implementing partners and countries have a shared understanding of what quality learning data means	Document review of partners' plans and reports ToC workshops with partners Global-level interviews with partners and FCDO lead Country-level interviews with government, partners and FCDO advisers
	4.2 To what extent are activities through the <i>Generate</i> strand likely to be able to meet the needs of both the <i>Track</i> and <i>Use</i> strands?	Partners delivering <i>Generate</i> strands have identified and prioritised needs of both global (<i>Track</i>) and national (<i>Use</i>) data users <i>Track</i> partners anticipate data generated through <i>Generate</i> activities being useful and usable National stakeholders anticipate data generated through <i>Generate</i> activities being useful and usable	Document review of partners' plans and reports ToC workshops with partners Global-level interviews with partners Country-level interviews with government, partners and FCDO advisers
5. To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies? (i.e. external coherence at country level)	5.1 To what extent are in-country programme activities aligned with national education structures and policies?	D4FL activities generate evidence to inform national education structures and policies The logistical features (e.g. timing, mode) of assessment activities established through D4FL align with national education structures	Document review of partners' plans and reports Country-level interviews with government, partners and FCDO advisers
	5.2 To what extent are the objectives and approaches of D4FL activities coherent within countries' assessment	D4FL activities support the right functions within countries' assessment systems to produce learning data that meets the needs of the <i>Track</i> and <i>Use</i> strands D4FL is not duplicating efforts	Country-level interviews with government, partners and FCDO advisers

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	systems (including assessment for learning, examinations, citizen-led assessments, sample-based standardised assessments)?		
6. To what extent are D4FL and other FCDO projects complementing one another? (i.e. internal FCDO coherence at country and global levels)	6.1 To what extent are D4FL and other FCDO projects actively engaging and complementing one another in-country?	D4FL and other projects are learning lessons from one another D4FL and other projects are aware of one another and how their activities fit together D4FL is providing guidance to FCDO education support programmes on generating and using learning outcome data	Document review of partners' plans and reports ToC workshops with partners Country-level interviews with government, partners and FCDO advisers
	6.2 To what extent are D4FL and other FCDO projects actively engaging and complementing one another at a global level?	D4FL and other projects are learning lessons from one another D4FL and other projects are aware of one another and how their programmes can fit together	Document review of FCDO documentation, guidance notes, D4FL annual reports Document review of partners' plans and reports Global-level interviews with partners and FCDO lead
<p>CRITERION: EFFICIENCY</p> <p>The extent to which D4FL is set up to and making best use of its resources to maximise the chance that programme outcomes will be achieved.</p>			
7. Will the programme set-up likely facilitate timely and efficient use of resources?	7.1 To what extent are current processes (contracting mechanisms, contractual arrangements and partners internal processes) likely to facilitate <u>timely</u> use of resources?	Processes are likely to facilitate timely responses to requests for support (within planned timeframes where applicable, and within timeframes that align with priorities for support and are acceptable to countries) Processes are likely to facilitate timely delivery of support (within timeframes that align with priorities for support where realistic, and at a pace that is acceptable countries and does not risk quality and relevance)	Document review of partners' plans Global-level interviews with partners and FCDO lead Country-level interviews with partners and FCDO advisers
	7.2 To what extent are current processes (contracting mechanisms, contractual	Processes are in place to ensure work by partners is well managed	Document review of partners' plans Global-level interviews with partners and FCDO lead

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	arrangements and partners internal processes) likely to facilitate <u>efficient</u> use of resources?	Processes in place are likely to minimise waste or duplication of efforts Processes are in place to ensure that resources are targeted at meeting programme outcomes and country priorities	Country-level interviews with partners
	7.3 To what extent and how are partners prioritising efficient use of resources?	Partner organisations have prioritised efficiency in their planning Partner organisations have put appropriate processes and protocols in place to ensure that resources are used efficiently (as defined above)	Document review of partners' plans Global-level interviews with partners and FCDO lead Country-level interviews with partners
8. To what extent are partners responding in a timely way to requests for support?	8.1 To what extent are partners responding in a timely way to initial requests for support?	Partners are responding to initial requests within planned timeframes (where applicable) Partners are responding to requests within timeframes that align with countries' priorities for the support Countries are satisfied with the timeliness of responses Countries remain engaged with the process	Document review of partners' plans and reports Country-level interviews with government, partners and FCDO advisers
	8.2 To what extent are partners working with countries to agree support packages in a timely way?	Partners agree support packages within planned timeframes (where applicable) Partners are agreeing support packages within timeframes that align with countries' priorities for the support Countries are satisfied with the pace of progress Countries remain engaged with the process Quality and relevance of support packages agreed are not impaired by a desire to move quickly	Document review of partners' reports Global-level interviews with partners Country-level interviews with government, partners and FCDO advisers
9. To what extent have D4FL resources been used efficiently to date, and its intended timeline kept?	9.1 To what extent have partners used D4FL resources efficiently to date?	Partners' work has been well project managed to result in quality outputs being delivered within expected level of effort Partners' work has involved minimal waste and duplication of efforts Partners have targeted resources at meeting programme outcomes and country priorities	Document review of partners' reports Global-level interviews with partners and FCDO lead Country-level interviews with partners and FCDO advisers

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	9.2 To what extent has FCDO's allocation and management of resources across strands and partners been efficient?	Minimal waste and duplication of efforts across strands and partners Resources have been used as planned	Document review of partners' reports Interview with FCDO lead
	9.3 Is the programme being delivered according to the intended timeline?	The programme timeline is being kept. If the programme timeline is not kept, the deviations are merited and likely to support better achievement of outcomes.	Interviews with implementing partners Interview with FCDO lead
CRITERION: EFFECTIVENESS The extent to which D4FL is likely to achieve intended results.			
10. To what extent are D4FL activities aligned with the programme ToC?	10.1 To what extent are plans and activities to date aligned with intended outcomes, including equity outcomes?	<i>Track</i> strand plans and activities to date are focused on producing learning data that: <ol style="list-style-type: none"> 1. is comparable 2. is disaggregated, including by gender and disability 3. includes emergency contexts <i>Generate</i> and <i>Use</i> strand plans and activities to date are focused on supporting LICs and MICs to generate data that is <ol style="list-style-type: none"> 4. robust 5. disaggregated, including by gender and disability 6. useful <i>Generate</i> and <i>Use</i> strand plans and activities to date are focused on supporting governments and teachers to use learning data to improve foundational learning	Document review of partners' plans and reports ToC workshops with partners Global-level interviews with partners Country-level interviews with government, partners and FCDO advisers
	10.2 To what extent are plans and activities under the generate and use strands aligned with the guiding principles?	Plans and activities to date are focused on providing support that is nationally and sub-nationally owned; and answers problems faced by countries, and effectively responds to their own demands and targets Plans and activities to date are focused on providing support that includes teachers and schools leaders and supports them to effectively do their job	Document review of partners' plans and reports ToC workshops with partners Global-level interviews with partners Country-level interviews with government, partners and FCDO advisers

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
		<p>Plans and activities to date are focused on providing support that is cost-effective and sustainable (i.e. not too burdensome for governments, schools and teachers), but allows for rigour, quality and comparability</p> <p>Plans and activities to date are focused on providing support that is effective at building state capability (not reliant on external or parallel processes)</p>	Interviews with external assessment experts
11. What has the programme achieved to date?	11.1 Are partners making expected progress in delivering activities and outputs?	<p>Implementing partners are making expected progress in delivering activities and outputs</p> <p>Outputs are being achieved within expected level of effort</p>	<p>Document review of partners' reports</p> <p>Global-level interviews with partners and FCDO lead</p> <p>Country-level interviews with government, partners and FCDO advisers</p>
	11.2 What results have been achieved to date regarding filling global learning data gaps?	More LICs and MICs are reporting internationally using comparable and disaggregated data on learning, including in emergency settings	<p>Document review of partners' reports</p> <p>Global-level interviews with partners</p>
	11.3 What results have been achieved to date regarding the generation of learning data in-country?	<p>Rigorous, regular primary learning assessments are being implemented</p> <p>Learning assessment is being embedded in national policies and reforms</p> <p>Simple foundational learning assessment is being used in classrooms and community interventions</p>	<p>Document review of partners' reports</p> <p>Global-level interviews with partners</p> <p>Country-level interviews with government, partners and FCDO advisers</p>
	11.4 What results have been achieved to date regarding the use of learning assessments to strengthen curriculum and instruction?	<p>Learning data is being used by government to implement and monitor interventions</p> <p>Learning data is being used by teachers and other educators to strengthen practice</p>	<p>Document review of partners' reports</p> <p>Global-level interviews with partners</p> <p>Country-level interviews with government, partners and FCDO advisers</p>
12. What are the likely barriers and facilitators to achieving results?	12.1 What are the likely barriers and facilitators to achieving results regarding filling global learning data gaps?	Assessment will be linked to emerging findings against earlier EQs and will include assessment of the validity of the key ToC assumptions	<p>Document review of partners' reports</p> <p>Interviews with external assessment experts</p> <p>Global-level interviews with partners and FCDO</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	12.2 What are the likely barriers and facilitators to achieving results regarding the generation of learning data in-country?	Assessment will be linked to emerging findings against earlier EQs and will include assessment of the validity of the key ToC assumptions	Document review of partners' reports ToC workshops with partners Interviews with external assessment experts Global-level interviews with partners and FCDO Country-level interviews with government, partners and FCDO advisers
	12.3 What are the likely barriers and facilitators to achieving results regarding the use of learning assessments to strengthen curriculum and instruction, including for marginalised groups?	Assessment will be linked to emerging findings against earlier EQs and will include assessment of the validity of the key ToC assumptions	Document review of partners' reports Interviews with external assessment experts ToC workshops with partners Global-level interviews with partners and FCDO Country-level interviews with government, partners and FCDO advisers
<p>LOOKING AHEAD TO IMPACT AND SUSTAINABILITY</p> <p>Articulating pathways and opportunities for sustained impact and influence to be inform the summative evaluation.</p>			
13. What are the likely pathways and timelines for outcomes and desired impact?	13.1 What are the likely timelines for the outcomes and desired impacts articulated in the D4FL ToC?	FCDO and partners' expected timelines for outcomes and impacts are realistic	Document review of literature on global assessment landscape and previous efforts Document review of partners' plans Interviews with external assessment experts ToC workshops with partners Country-level interviews with partners
	13.2 What are the expected pathways to intended outcomes, and therefore what early indicators of promise should the	Pathways in the final explicated Theory of Change for testing in the summative evaluation, are clear and credible	Document review of literature on global assessment landscape and previous efforts Document review of partners' plans

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	summative evaluation aim to capture?	Summative evaluation can provide evidence of expected pathways and early indicators of promise within the duration of the evaluation	Interviews with external assessment experts ToC workshops with partners Country-level interviews with partners
14. What are the avenues for D4FL to have sustainable impact?	14.1 How do FCDO and partners conceptualise sustainability for D4FL?	Clear and credible conceptualisation of how outcomes achieved through the D4FL, will be sustained Conceptualisations of sustainable impact are clear and credible Conceptualisations of sustainable impact align with countries' and global sector's needs FCDO and partners' conceptualisations are aligned	Document review of literature on global assessment landscape and previous efforts Interviews with external assessment experts Document review of partners' plans ToC workshops with partners Global-level interviews with partners and FCDO lead Country-level interviews with partners and FCDO advisers
	14.2 To what extent and how has sustainability been integrated into programme design?	Sustainability has been appropriately integrated into programme design	Document review of partners' plans ToC workshops with partners Global-level interviews with partners and FCDO lead
	14.3 What other opportunities are there for sustainable impact?	n/a	Document review of literature on global assessment landscape and previous efforts Interviews with external assessment experts
15. What are the avenues for FCDO to influence the wider sector?	15.1 How does FCDO hope to influence the wider sector?	Aspirations for influence clear and credible Aspirations for influence align with global sector's needs FCDO and partners' conceptualisations are aligned	Document review of literature on global assessment landscape and previous efforts Interviews with external assessment experts Interview with FCDO lead

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
			Country-level interviews with FCDO advisers
	15.2 How can learning and outputs from D4FL benefit the sector?	n/a	<p>Document review of literature on global assessment landscape and previous efforts</p> <p>Interviews with external assessment experts</p> <p>Global-level interviews with partners and FCDO lead</p> <p>Country-level interviews with partners and FCDO advisers</p>

Annex 4 Original & Updated Theory of Change, Partner Theories of Change

In this Annex, Figure 1 sets out the original D4FL Programme Theory of Change (ToC). Please refer to Section 3.2 of this report for an explanation of the programme design and theory to accompany diagrammatic ToC.

Figure 2 sets out an updated ToC for discussion. As explained in Section 5.2, we recommend to adjust the ToC to reflect current programme design and intended mechanisms of change. An accurate ToC will enable the summative evaluation to accurately evaluate the programme against its intended outputs and outcomes. In addition, the ToC can serve as a programming tool, helping partners understand how the different strands of work fit together, and enabling learning conversations about how well the different strands are progressing and how they might better join up.

We have proposed adjustments to the ToC to reflect all the currently intended D4FL inputs, outputs, intermediate outcomes, outcomes and impact, and how they are expected to lead to one another. Specifically, we have:

- set out three programme outcomes, reflecting the aims of the three programme strands (*Track, Generate and Use*)
- rephrased the (now) third outcome so that capability to use data is not limited to state actors
- shown how these three programme outcomes are expected to contribute to the ultimate impact of improved foundational learning, through three pathways (increased pressure for action on foundational learning, when evidence is available of low learning levels; and increased quantity and quality of action on foundational learning as a result of the availability of data and state and non-state capability to use it)
- depicted how both the increased pressure for action on foundational learning, and the “proof of concept” provided by the programme are expected to lead to increased demand for data in countries and among actors other than those directly supported by the programme
- separated inputs so that it is clear which outputs they lead to, and removed that which will not be funded (i.e. PAL Network work on action in the classroom)
- converted all outputs to programme outputs rather than FCDO actions (e.g. where outputs previously stated “fund...”)
- added an output around strengthening curricula, teacher, development and teaching materials and shown how this is expected to directly lead to the ultimate outcome of improved foundational learning, outside of the data-focused mechanisms set out in the ToC
- added assumptions articulated by programme stakeholders in the formative evaluation, and indicated where in the TOC they apply

We have also greyed out the one programme input that was planned but to date has not transpired.

In the following figures 3-6, we set out a tailored version of the ToC for each currently implementing partner (and separately for the World Bank’s ALMA and Accelerator), indicating which inputs, outputs, outcomes, causal mechanisms and assumptions are relevant for that piece of the programme.

At the start of the mid-term evaluation in 2025, the proposed TOCs will be reviewed with the D4FL Programme Team and partners, to set the frame for this evaluation round.

Figure 1 Original D4FL Programme Theory of Change

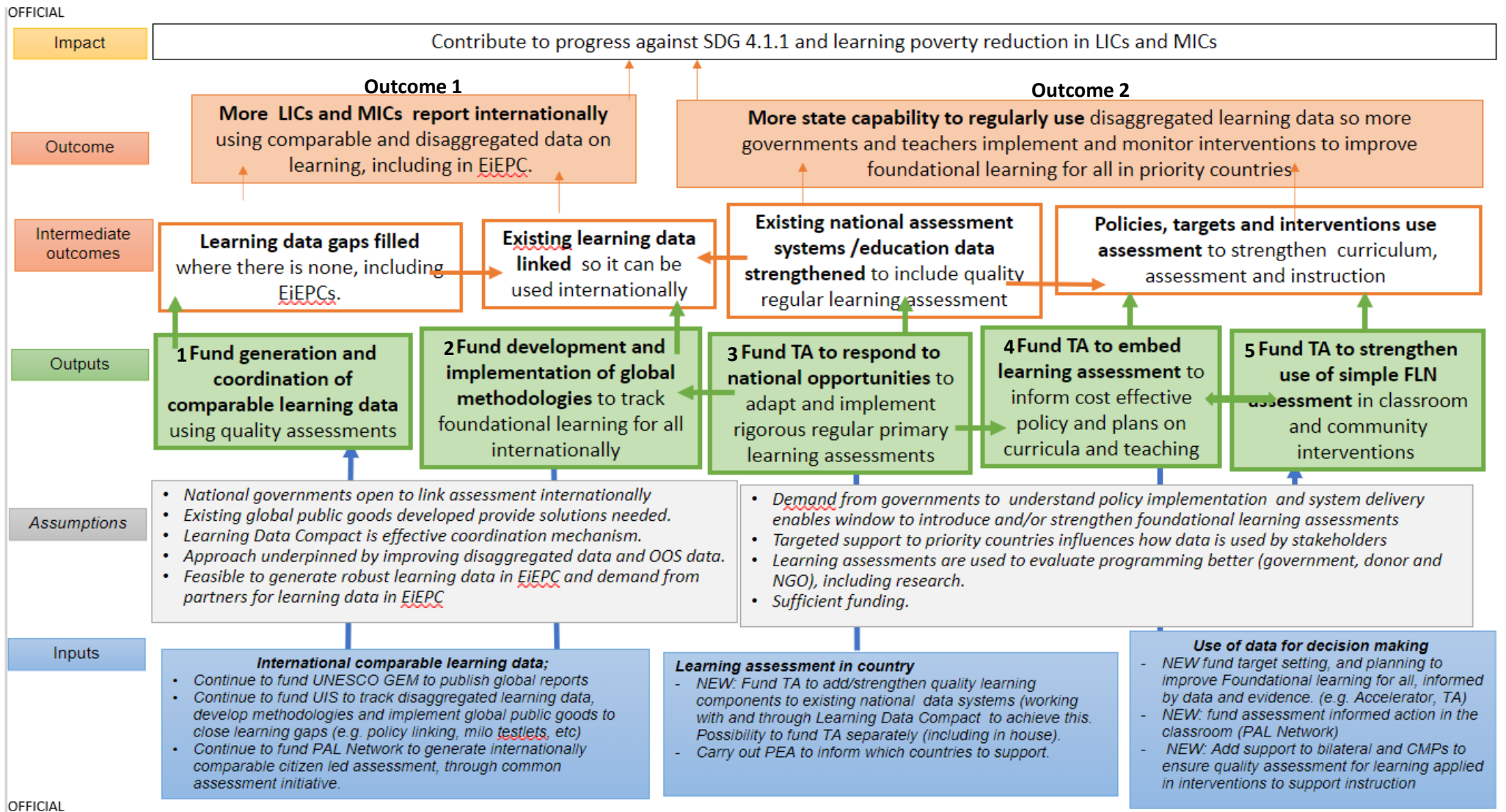
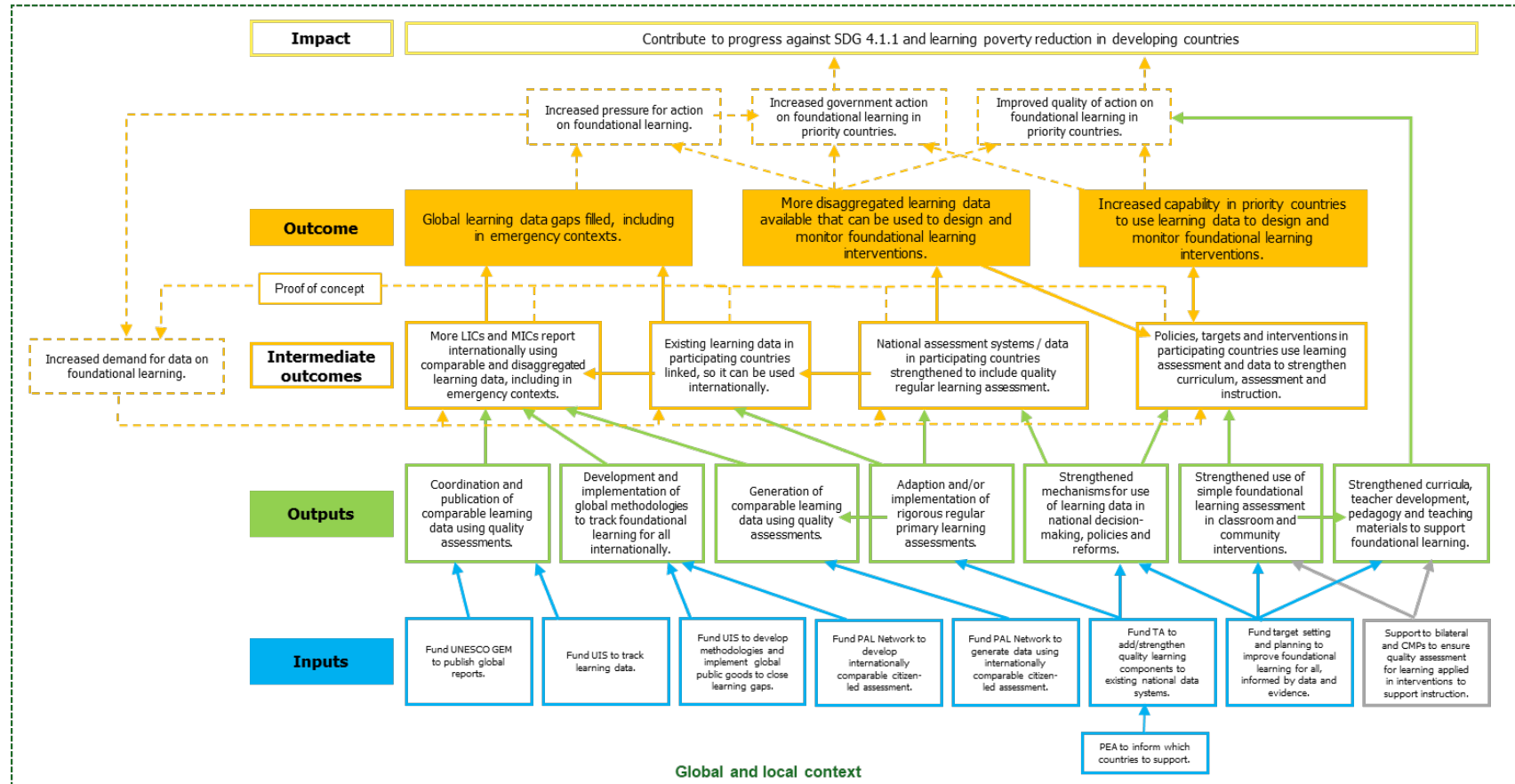


Figure 2 Updated D4FL Theory of Change



Assumptions

1. Sufficient funding is available.
2. Learning Data Compact and other coordination mechanisms are effective.
3. National governments are open to link assessment internationally.
4. There is demand from governments to understand policy implementation and system delivery enables window to introduce and/or strengthen foundational learning assessments.
5. Targeted support to priority countries influences how data is used.

Inputs to outputs

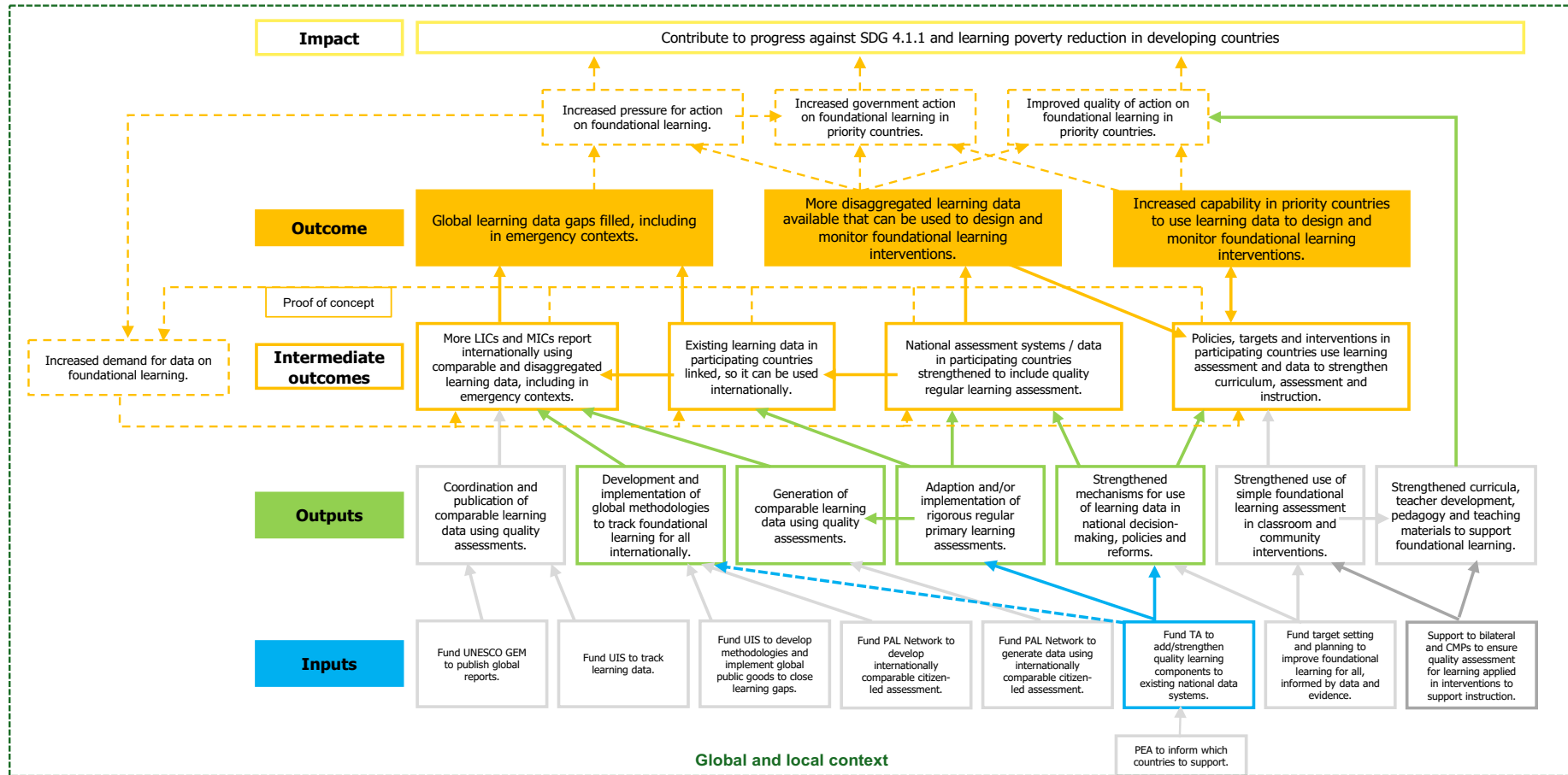
Outputs to outcomes

6. Global public goods developed provide solutions needed.
7. National governments are open to link assessment internationally.
8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises.
9. Approach is underpinned by improving disaggregated data.
10. The approach is underpinned by improving out-of-school data.
11. It is feasible to generate robust learning data in emergency contexts.
12. There is demand from partners for learning data in emergency contexts.
13. Learning assessments are used to evaluate programming better.
14. Requirements for international reporting are clear and consistent.

Outcomes to impact

15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundation learning.
16. Governments and/or civil society have the resources (financial, human, technical) to (continue to) carry out learning assessment/data generation exercises.
17. Governments respond to increased pressure for action on foundational learning.

Figure 3 ACER UK ToC mapped onto D4FL ToC



Inputs to outputs

Outputs to outcomes

Assumptions

1. Sufficient funding is available.
 2. Learning Data Compact and other coordination mechanisms are effective.
 3. National governments are open to link assessment internationally.
 4. There is demand from governments to understand policy implementation and system delivery enables window to introduce and/or strengthen foundational learning assessments.
 5. Targeted support to priority countries influences how data is used.
- Outputs to outcomes**
6. Global public goods developed provide solutions needed.
 7. National governments are open to link assessment internationally.

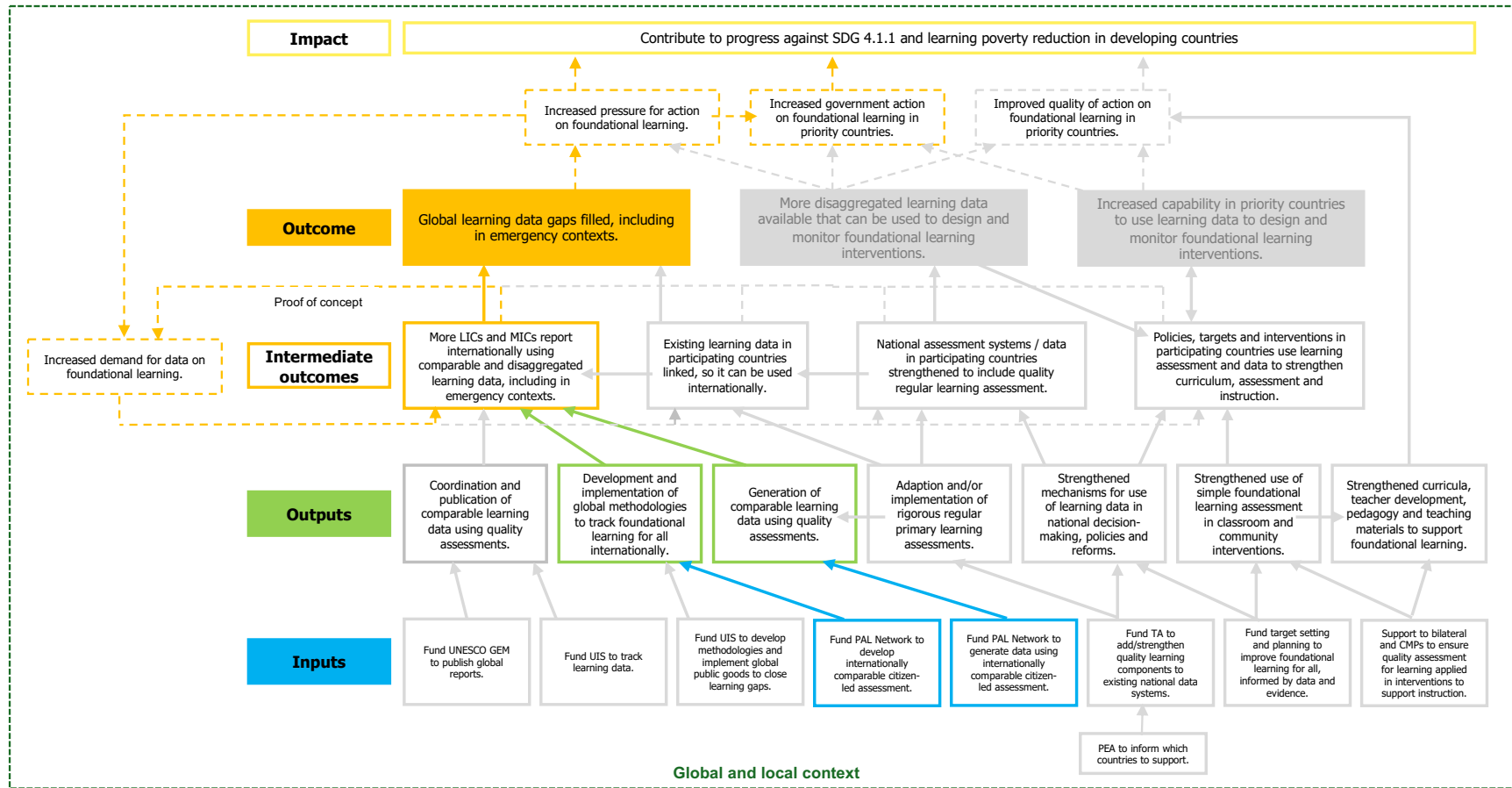
8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises.
9. Approach is underpinned by improving disaggregated data.
10. The approach is underpinned by improving out-of-school data.
11. It is feasible to generate robust learning data in emergency contexts.
12. There is demand from partners for learning data in emergency contexts.
13. Learning assessments are used to evaluate programming better.
14. Requirements for international reporting are clear and consistent.
15. There is a culture of transparency.
16. UIS criteria are not changed and the process for approving evidence is agreed.

17. Suitable level of in-country coordination and policy consistency, to ensure that work is not in conflict with other country priorities and activities.
18. Coordination and communication across partners stays positive and there is an effort to avoid duplication.
19. The new approach is scalable and meets different needs/purposes of Government.
20. Political change will not affect agreed priorities on foundational learning assessments.

Outcomes to impact

21. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundation learning.
22. Governments and/or civil society have the resources (financial, human, technical) to (continue to) carry out learning assessment/data generation exercises.
23. Governments respond to increased pressure for action on foundational learning.

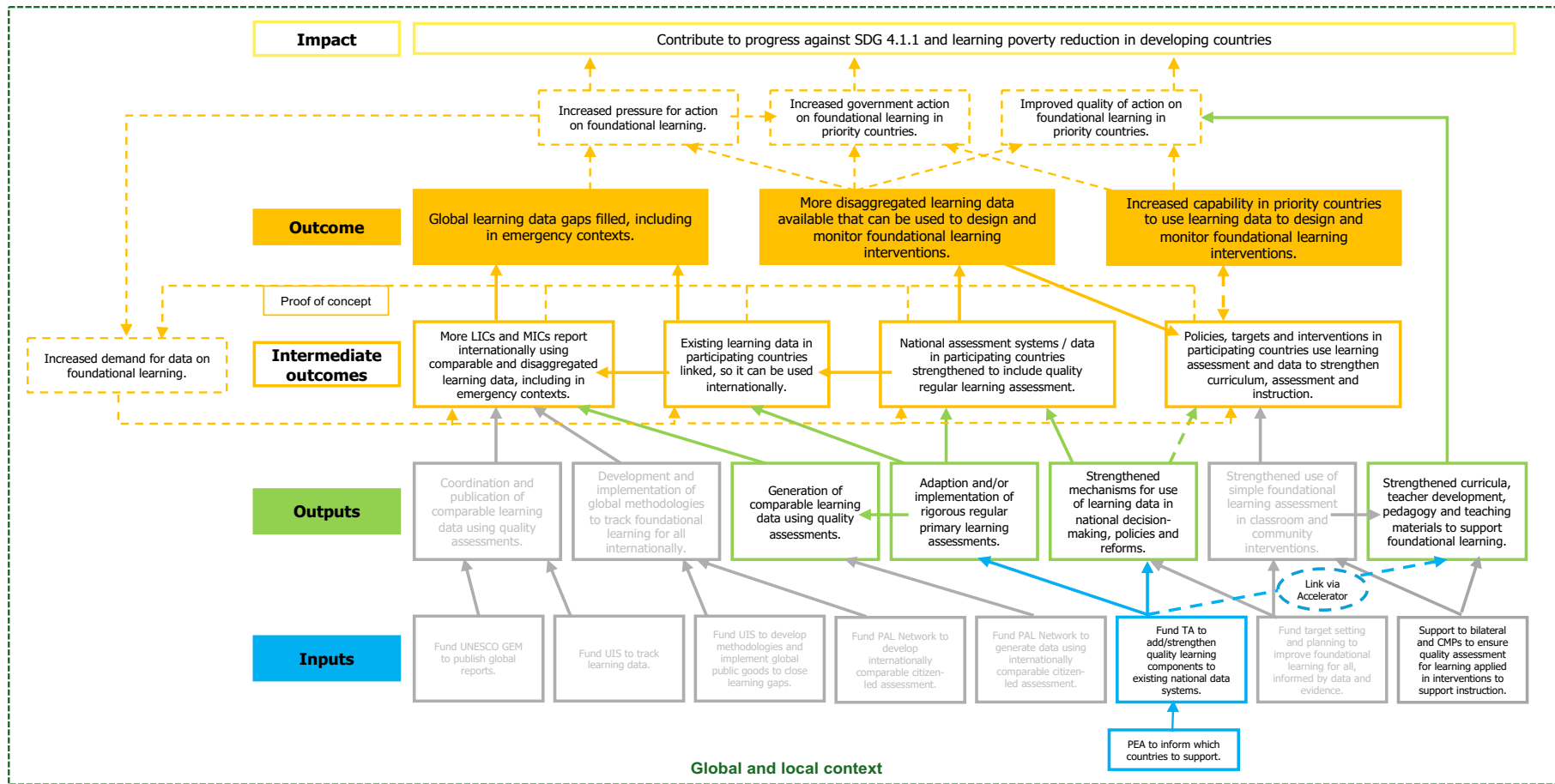
Figure 4 PAL Network ToC mapped onto D4FL ToC



Assumptions

- | Inputs to outputs | Outputs to outcomes | Outcomes to impact |
|--|--|---|
| <ol style="list-style-type: none"> 1. Sufficient funding is available. 2. Learning Data Compact and other coordination mechanisms are effective. 3. National governments are open to link assessment internationally. 4. There is demand from governments to understand policy implementation and system delivery enables window to introduce and/or strengthen foundational learning assessments. 5. Targeted support to priority countries influences how data is used. 6. Strong and sustained involvement from local communities and partners. 7. Government agencies provide necessary authorisations and support for sampling and data generation activities. 8. Implementing partners have adequate technical capacity to ensure quality delivery. 9. Political will/stability?? | <ol style="list-style-type: none"> 6. Global public goods developed provide solutions needed. 7. National governments are open to link/report assessment internationally. 8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises. 9. Approach is underpinned by improving disaggregated data. 10. The approach is underpinned by improving out-of-school data. 11. It is feasible to generate robust learning data in emergency contexts. 12. There is demand from partners for learning data in emergency contexts. 13. Learning assessments are used to evaluate programming better. 14. Requirements for international reporting are clear and consistent. 15. Funded tools and methodology are recognised by UIS. | <ol style="list-style-type: none"> 15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundation learning. 16. Governments and/or civil society have the resources (financial, human, technical) to (continue to) carry out learning assessment/data generation exercises. 17. Governments respond to increased pressure for action on foundational learning. 18. Generated evidence is turned into quality and accessible awareness and advocacy products. 19. Dissemination and policy advocacy efforts are sufficient to increase the salience of foundational learning. |

Figure 5 World Bank ALMA ToC Mapped onto D4FL ToC



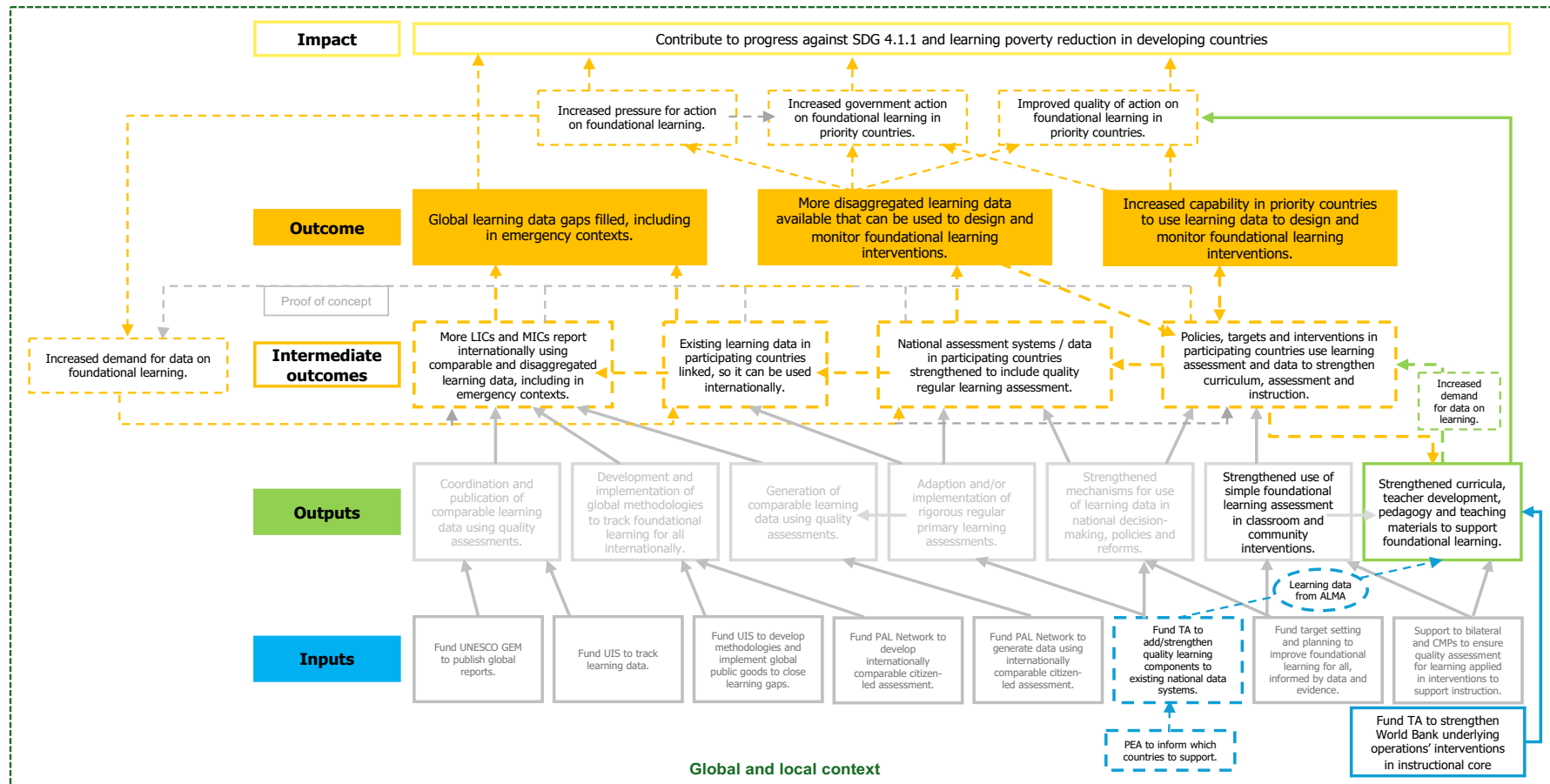
Assumptions

- Inputs to outputs**
1. The funding is sufficient for the TA to deliver quality outputs .
 2. Learning Data Compact and other coordination mechanisms are effective.
 3. National governments are open to link assessment internationally.
 4. Demand from governments to understand policy implementation and system delivery enables windows to introduce and/or strengthen foundational learning assessments.
 5. Country Teams / the underlying World Bank operation identified the most urgent complementary TA needed to strengthen mechanisms
 6. World Bank regional Practice Managers and Country Teams use ALMA learning data to identify what instruction core interventions require support (also in Accelerator where it overlaps
 7. Targeted support to priority countries influences how data is used.

- Outputs to outcomes**
6. Global public goods developed provide solutions needed.
 7. National governments are open to link assessment internationally.
 8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises.
 9. Approach is underpinned by improving disaggregated data.
 10. The approach is underpinned by improving out-of-school data.
 11. It is feasible to generate robust learning data in emergency contexts.
 12. There is demand from partners for learning data in emergency contexts.
 13. Learning assessments are used to evaluate programming better.
 14. Requirements for international reporting are clear and consistent.

- Outcomes to impact**
15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundation learning.
 16. Governments and/or civil society have the resources (financial, human, technical) to (continue to) carry out learning assessment/data generation exercises.
 17. Governments respond to increased pressure for action on foundational learning.

Figure 6 World Bank Accelerator ToC mapped onto D4FL ToC



Assumptions

Inputs to outputs

Outputs to outcomes

Outcomes to impact

1. The funding is sufficient for the TA to deliver quality outputs .
2. Learning Data Compact and other coordination mechanisms are effective.
3. National governments are open to link assessment internationally.
4. Demand from governments to understand policy implementation and system delivery enables windows to introduce and/or strengthen foundational learning assessments.
5. Country Teams / the underlying World Bank operation identified the most urgent complementary TA needed for effective interventions in instructional core
6. World Bank regional Practice Managers and Country Teams use ALMA (where in place) or other learning data to identify what instruction core interventions require support
7. Targeted support to priority countries influences how data is used.

6. Instructional core interventions stimulate need for more data
7. Global public goods developed provide solutions needed.
8. National governments are open to link assessment internationally.
9. Governments and/or civil society have the resources (financial, human, technical) to carry out and sustain instructional core interventions, learning assessment/data generation exercises.
10. Ability of countries to use data is enabled by improved disaggregated data.
11. Countries instructional core interventions include out of school children and children with disabilities.
12. It is feasible to generate robust interventions and learning data in emergency contexts.
13. There is demand from partners for learning data in emergency contexts.
14. Learning assessments are used to evaluate programming better.
15. Requirements for international reporting are clear and consistent.

15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundation learning.
16. Governments and/or civil society have the resources (financial, human, technical) to (continue to) carry out learning assessment/data generation exercises.
17. Governments respond to increased pressure for action on foundational learning.

Annex 5 Use and Influence Plan

42. This annex outlines the progress made against the Use and Influence Plan defined in the IR, specifically within Sections 4.5: Audience Analysis and Dissemination, and 4.6: Communication Strategy. It is structured into three parts:

- A restatement of the principles, outputs, and engagement strategies as articulated in the IR.
- The Use and Influence Plan, including planned dissemination outputs and engagement activities.
- A systematic review of progress to date, highlighting what has been achieved, areas requiring further action, and how we plan to use findings from the Formative ER to meet our goals.

The Use and Influence Plan

43. As described in the IR, the Use and Influence Plan adopts a strategic and multifaceted approach to ensure the effective uptake of evaluation findings by primary and secondary audiences. Key principles, planned outputs, and active engagement strategies are summarized below:

Principles for Evidence Uptake

44. **Variety of Materials:** Employing a pyramid approach to create outputs tailored for different levels of engagement and expertise.

45. **Accessibility:** Ensuring outputs are clear, minimise technical jargon, and include simple summaries where needed.

46. **Active Engagement:** Combining written outputs with workshops, webinars, and interactive sessions to maximize evidence use.

Outputs

47. The written outputs for the Formative Evaluation include:

- Explicated Theory of Change
- Formative Evaluation Report
- Evaluation Digest
- Summative Evaluation Plan

Active Engagement

48. Engagement activities to disseminate findings included:

- Learning and Reflection Workshop with FCDO
- Learning and Reflection Workshop with D4FL Partners

Communications Strategy

49. Communication has been designed to ensure clarity and alignment with stakeholders at every stage. This includes:

- Regular communication with FCDO through quarterly reports and meetings.
- Engagement with implementing partners, including updates and collaborative planning for dissemination.

Progress to date and future steps

Outputs Delivered

50. Explicated ToC: Completed and shared with the Steering Committee for feedback in the FCDO workshop.

51. Learning and Reflection Workshops: Both workshops with FCDO and implementing partners were successfully conducted, with key insights incorporated into the formative evaluation report.

Outputs Underway

52. Formative Evaluation Report: Drafting is on schedule.

53. Evaluation Digest: A draft structure has been developed, with content currently being finalised based on the Formative Evaluation Report, to be submitted mid-January 2025 with the 2nd draft Formative Evaluation Report.

54. Summative Evaluation Plan: A draft will be developed upon completion of the Final Formative Evaluation Report.

Engagement Activities

55. Active engagement sessions for both the Steering Committee and implementing partners were well-received, fostering interactive discussions and co-development of actionable insights.

56. Feedback from workshops is directly informing the refinement of outputs, ensuring alignment with audience needs.

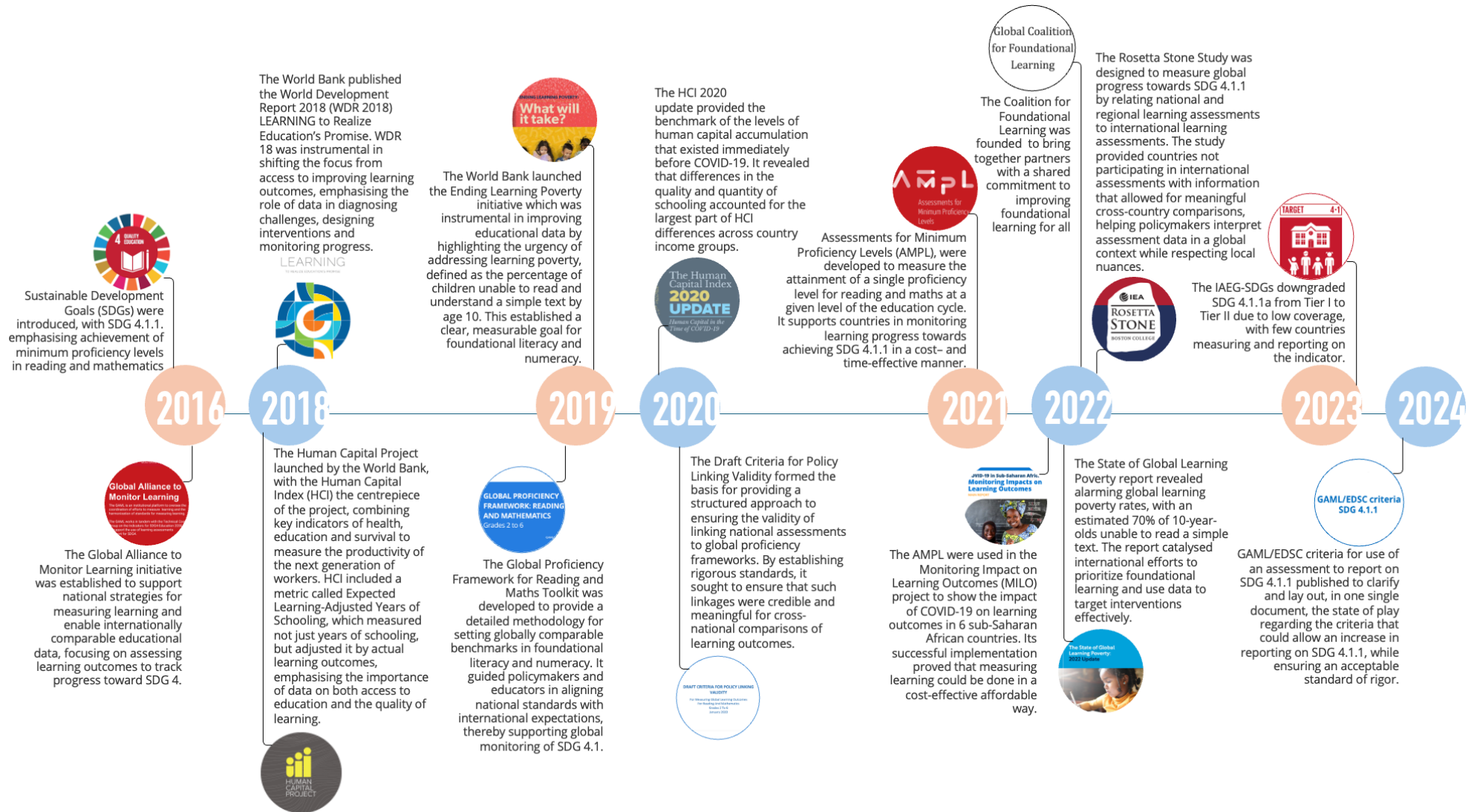
Remaining Tasks

57. Finalize the Formative Evaluation Report and accompanying Evaluation Digest by 31 January 2025.

58. Initiate additional engagement activities to support the summative evaluation design, as per workshop outcomes.

Annex 6 Timeline of Events in the Data for Foundational Learning Context

Figure 7 Timeline of events



Annex 7 Context for D4FL: Learning assessment in the context of SDG4

Political and technical caveats of assessment of learning outcomes for SDG4

59. SDG 4 aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (UNESCO, 2015). It marks the first truly global agreement in education. For the first time, countries from diverse economic and cultural backgrounds agreed on a common set of educational objectives, demonstrating a collective commitment to improving education worldwide (UNESCO, 2017). This agreement was a significant milestone, given the varied priorities and challenges faced by different nations.

1. One of the most notable changes in SDG4 in comparison to previous education goals is its shift in focus from access to quality in education (Sayed & Moriarty, 2020). Unlike former global education initiatives that primarily aimed to increase access, measuring school availability, enrolment, completion, etc., SDG4 emphasizes the importance of ensuring that students are actually learning once they are in the classroom. By doing this, SDG4 acknowledges a critical reality: merely attending school does not guarantee that children are actually learning (Addey, 2024). While this focus on quality is a significant advancement, it also presents the challenge of defining what quality education entails. Consequently, one of the major outcomes of this shift has been the heightened importance of assessing learning outcomes. This need to measure educational quality highlights the need for robust and reliable assessment systems to evaluate whether students are achieving the intended learning objectives.

2. While often presented as something purely technical, educational assessment is also deeply political. What we choose to assess signals what we consider important, and thus inherently reflects our priorities. For example, teachers are often advised that good assessments should reflect the distribution of instructional time they used in their classrooms. Similarly, setting national, regional, or global targets involves reflecting on social priorities. The methods, criteria, scoring mechanisms and dissemination of results involved in an assessment will reflect broader worldviews, cultural norms, priorities, political will, and power dynamics.

3. This political dimension was evident in the development of SDG4. During the discussions leading to the agreement, tensions arose between stakeholders with differing agendas. Some proponents, such as the OECD, advocated for a stronger focus on an economics-driven vision of education that highlighted its capacity to help people find better work and improve their standards of living, while contributing to more solid economies. Others, such as UNESCO, were more closely aligned with a human rights approach that understands education as a right that enables other rights, with a mission of helping learners to develop holistically and be able to coexist with one another, contributing to stronger democracies and peaceful societies (Addey, 2024). As SDG4 was to reflect global priorities in education, it attempted to capture both perspectives. While it emphasized quality and inclusivity as an overarching goal, it included specific targets to address each view, such as 4.4, which refers to relevant skills for employment, and 4.7, which refers to holistic skills to promote global citizenship (UNESCO, 2015).

4. Specific targets, such as those mentioned above, come from the need for operationalizing SDG4 into measurable outcomes, so that it can be monitored and assessed, ensuring accountability. SDG4 has been broken down into ten targets, each with associated indicators. However, operationalizing such a comprehensive goal and deciding on valid metrics for each of its aspects is not straightforward. Particularly as the focus shifts from concrete measures like school enrolment to more abstract and contested concepts like educational quality, there is no one way to define progress

(Sayed & Moriarty, 2020). In the case of SDG4, the indicators chosen were influenced both by political aspects and by data availability. For example, target 4.1 states “By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes” (UNESCO, 2015). While concepts of equity, quality, relevance and effectiveness can be interpreted quite broadly, the indicators for this target refer only to proficiency levels in math and reading, and completion rates in formal schooling. Because concrete targets drive efforts towards measurable outcomes, often summarized in the principle “what gets measured gets done,” it is likely that these indicators will guide focus and resources towards these important, but not all-encompassing aspects of learning, and away from the holistic human rights approach that seems to guide the overall goal.

5. These subtle shifts happen for several reasons. First, some concepts are inherently more challenging to measure. For instance, assessing a child's ability to add is easier than evaluating its empathy. This often leads to a focus on aspects with clearer indicators that are less contentious and where showcasing progress is simpler. An example of this is target 4.7, which ended up capturing most of the holistic and more controversial aspects of SDG4 under a very vague indicator,¹⁷ that makes tracking and accountability complicated, and thus the target easier to overlook (Sayed & Moriarty, 2020). Additionally, stakeholders have different views on what should be assessed and how, and varying levels of influence in pushing for their preferred indicators. Education ministries might have different priorities than teachers, for example, but even similar stakeholders can have differing worldviews, such as the tensions mentioned above between the OECD and UNESCO (Addey, 2024).

6. Another challenge is defining proficiency levels. The Global Alliance for Measuring Learning Assessment (GAMLA) has made significant efforts to reach a global agreement on this issue by holding international discussions and creating frameworks that fit different educational contexts and allow for comparison between countries. Though some progress has been made with the development of benchmarks that attempt to balance rigor with inclusivity, obstacles remain. Because of vastly different educational realities, including systemic aspects, curricula, resource availability and teacher training, learning outcomes considered minimal in the global North may still be unattainable in many LICs and MICs. At the same time, governments might resist setting proficiency levels that would highlight widespread failure to meet targets, pushing to establish lower benchmarks. These disparities often result in benchmarks that may not fully reflect the educational realities of all countries, complicating efforts to achieve a fair and accurate assessment of global learning.

7. Lastly, there are tensions on the decisions about which assessments to use to inform the monitoring of proficiency levels, even at the national level. Ideally, this decision should respond to a clearly defined purpose, and use methods that are aligned with it, as well as with resource availability. In terms of purpose, assessments can be broadly classified as formative (for learning) and summative (of learning) assessments. If the main purpose of an assessment is to provide feedback to actors, it should avoid tying results to consequences, to discourage practices such as teaching to the test, instead incentivising responses that reflect the real ability of those assessed as closely as possible. In contrast, if the main purpose of an assessment is to ensure accountability, especially if there are consequences associated to the results, the focus should be on rigour and comparability, but it should be noted that the results are less likely to provide the best possible feedback to teachers and learners. There will be more incentives for participants to cheat, opt out, or be excluded, in an attempt to avoid negative or attain positive consequences.

¹⁷ Extent to which (i) global citizenship education and (ii) education for sustainable development are mainstreamed in (a) national education policies; (b) curricula; (c) teacher education and (d) student assessment.

8. In terms of methods, both summative and formative assessments can use a myriad of formats: they can be implemented in a group context or individually, analyse processes more deeply for example through portfolio work or opt for faster and easier to grade approaches such as multiple-choice exams, create rubrics to provide more flexibility when evaluating a task, etc. Some formats might be better suited for a particular purpose, for example, standardised assessment is often more useful for summative assessment, but this will largely depend on the design of the actual tools. Several countries use portfolio work as end of school examinations or to determine college entrance, and multiple-choice exams can be used as formative assessment by using results to provide feedback to students.

9. Despite these important differences, there is often a mismatch between assessment and purpose. While both formative and summative assessments are crucial and do not depend on a single format, International Large Scale Assessments (ILSAs) are often perceived as the gold standard due to their focus on standardisation and statistical rigour, leading to some of their characteristics unnecessarily permeating evaluations that have different goals.

10. There is a rich body of literature on ILSAs, their background, aims, objectives, use and application of data as well as various reports on national performance and cross-country comparisons (Addey, Sellar, Steiner-Khamsi, Lingard, & Verger, 2017).

11. The perspective that ILSAs represent the gold standard in international assessment is increasingly questioned, and measurement approaches are shifting to nimbler, more cost-effective approaches. ILSAs are expensive, time-consuming, complex (Wagner, Wolf, & Boruch, 2018) and require significant technical capacity and expertise. ILSAs provide a single snapshot in time as they are conducted at a certain frequency, and the results are often not released for years after data collection. For this reason, their use as “whips” or “thermometers” (Lockheed & Wagemaker, 2013) able to inform educational policy has been questioned. This has led researchers to call for “smaller, quicker and cheaper” tools that are aligned with the policy goals desired, giving due consideration to timing and resourcing constraints (Wagner, Wolf, & Boruch, 2018).

12. No one assessment can serve all purposes and misalignment of assessment approach and purpose generates resistance from stakeholders. For example, there is a tendency to use results of supposedly formative assessments for accountability purposes and use national examination results to monitor national progress, despite them being less representative of the national reality.

13. Overall, we can see that while there is a formal consensus around SDG4, its implementation has revealed more nuanced challenges that need to be addressed. Countries have agreed to the 2030 deadline, but there are still no clear accountability mechanisms to ensure adherence, and debates about the methods, timing, and consequences of failing to reach targets remain. Countries' willingness to align with certain perspectives and adapt their policies might often depend more on their varying dependencies on international organizations' influence than on their own commitment to the SDG4 targets and their trust in assessment outcomes. Ensuring uptake, such that findings from SDG monitoring translate into political action and educational policy and practice, remains a significant challenge.

Progress towards foundational learning according to current SDG4 monitoring

14. Of the 10 targets and 12 indicators, five targets and six indicators directly relate to learning outcomes or skills. This captures the shift towards the quality of education that distinguishes SDG4 from previous global education goals. Among these, target 4.1 is particularly relevant for primary education. It aims to ensure that all girls and boys complete free, equitable, and quality primary and secondary education leading to relevant and effective learning outcomes. The most pertinent

indicator for this target for primary education is 4.1.1, which measures the proportion of children achieving at least a MPL in reading and mathematics at grade 2/3 and by the end of primary education (UNESCO, 2015). Math and reading were chosen as subjects because they are considered key within foundational learning, and metrics are more readily available than for other transferable skills.

Data sources and availability for SDG 4.1.1

15. Indicator 4.1.1 required defining what the MPL would be, which posed significant challenges. Curricula vary widely across countries, and cultural differences influence what is considered relevant. To establish a global MPL, several existing assessments were analysed and compared to identify common performance level descriptors and be used as guidance during GAML consensus building meeting in 2018 (UNESCO, 2018). This effort resulted in the global descriptor for MPL outlined in Section 3.1.

Table 12 Primary education MPL and equivalencies with cross-national assessments

Domain	MPL		Equivalencies in cross-national assessments		
	2/3 Grade	End of primary	Assessment	Age/Grade	Level considered MPL
Math	Students demonstrate skills in number sense and computation, reading simple data displays, shape recognition and spatial orientation.	Students demonstrate skills in number sense, computation, real world problems, basic measurement, 2D shape recognition, and reading and interpreting simple data displays.	ASER	6- to 14-y/o	Not applicable
			EGMA	Grade 1 to 3	Not applicable
			ERCE	Grade 3	Level 2
				Grade 6	Level 3
			PASEC	Grade 2	Level 2
				Grade 6	Level 3
			PILNA	Grade 4	Not applicable
				Grade 6	Level 6
			SACMEQ	Grade 6	Level 5
			SEA - PLM	Grade 5	Level 6
			TIMSS	Grade 4	Intermediate
UNICEF MICS 6	5- to 17-y/o	Not applicable			
UWEZO	6- to 16-y/o	Not applicable			
Reading	Students read aloud and comprehend many single written words, particularly familiar ones, and extract explicit information from sentences. They make simple inferences when longer texts are read aloud to them	Students independently and fluently read simple, short narrative and expository texts. They locate explicitly stated information, interpret and give some explanations about the key ideas in these texts. They provide simple, personal opinions or judgements about the information, events and characters in a text.	ASER	6- to 14-y/o	Not applicable
			EGRA	Grade 1 to 3	Not applicable
			ERCE	Grade 3	Level 2
				Grade 6	Level 3
			PASEC	Grade 2	Level 3
				Grade 6	Level 4
			PILNA	Grade 4	Not applicable
				Grade 6	Level 5
			SACMEQ	Grade 6	Level 5
			SEA - PLM	Grade 5	Level 6
			PIRLS	Grade 4	Low
UNICEF MICS 6	5- to 17-y/o	Not applicable			

16. Despite these definitions, some challenges persist. First, not all countries have collected learning outcome data, and those that have, have not necessarily done so at the target's grade levels. Even where data exists, aligning it with the global descriptor is complex. Much of the existing data was gathered before the MPL was agreed, or without the intent to report towards it being the main goal,

leading to potential misalignments. For example, national assessments might prioritise monitoring progress towards curriculum, choosing to measure skills different from those in the MPL. In other cases, it has been observed that some assessments highest possible performance level is still lower than the MPL, making them ineffectual for monitoring purposes. These challenges require compromises between an alignment with the MPL that allows for comparability, and data collection expectations that are feasible for all countries. Responding to this need, the UNESCO Institute for Statistics (UIS) has developed guidance for countries on using available data, prioritizing it, and linking it to international standards, yet issues of data reliability and credibility of these linking exercises remain (UNESCO, 2022).

17. According to the latest available data, approximately 70% of education systems have reported some data towards indicator 4.1.1. However, only about 53% of education systems have enough data to be able to report trends. It should also be noted that these percentages vary widely across regions: as shown in Table 13. Thus, while 78% of education systems in Europe and Northern America can report trends in indicator 4.1.1, only 10% can in Oceania. All in all, around half of the world's children currently live in countries without available data to report trends in indicator 4.1.1. The availability of data tends to increase with grade level, meaning that, in general, the largest gaps are for learning outcomes at grades 2 and 3, while the largest availability is for children in lower secondary (Gustaffson, 2022).

Table 13 Availability of data to report against indicator 4.1.1 across regions

Region	Education systems (including territorial entities and regions)	Countries with 4.1.1 statistics	% of countries with 4.1.1 statistics	Countries with 4.1.1 trends	% of countries with 4.1.1 trends
Sub-Saharan Africa	52	40	77%	29	56%
Northern Africa and Western Asia	24	21	88%	20	83%
Eastern and South-eastern Asia	18	16	89%	8	44%
Central and Southern Asia	14	11	79%	4	29%
Oceania	21	6	29%	2	10%
Latin America and the Caribbean	48	24	50%	17	35%
Europe and Northern America	51	42	82%	40	78%
World	228	160	70%	120	53%

Source: Adapted from Gustaffson, 2022

18. At the same time the quality, reliability and alignment with MPL of the data used for reporting is very different across education systems. Countries use a variety of sources and are instructed to prioritise data from international assessments like PISA and TIMSS, then data from regional assessments such as SACMEQ and PASEC, and finally national assessment data. However, some of the most common international assessments, such as EGRA and EGMA, as well as several regional assessments, were considered unsuitable for reporting purposes for different reasons, including representativity, administration method and comparability (Montoya, 2024). Data coming from national assessments often suffers from similar issues.

Progress towards SDG 4.1.1 according to available data

19. Regarding progress, most countries are not on track to achieve SDG4 by the target year of 2030. According to estimates of the IAEG-SDG in 2023, even if trends reported before the pandemic remained unchanged, only 1 in 6 countries would meet the goal. For indicator 4.1.1 it was projected that 300 million students would not have the necessary skills in math and reading they needed to succeed in life (United Nations Economic and Social Council, 2023). This is not to say that no progress has been made. Given the complex nature of the data, it is hard to estimate trends at a global level, but reports attempting to extrapolate findings from ILSA such as PIRLS had found modest improvements of around 0.7 percentage points a year, that are consistent with available datapoints at the national level (Gustaffson, 2022). However, these advances have been slowing down consistently, leading to a diagnosis of stagnation in progress towards indicator 4.1.1.

20. In addition to this, the COVID-19 pandemic has had a profound impact. Education systems worldwide were disrupted, and often forced to fully shut down schools for an average of 92 days globally but reaching even more in some regions, with the highest being Latin America and the Caribbean with an average of 158 days of school closures (UNICEF, 2021). The pandemic generated a learning crisis and exacerbated existing disparities. Though countries made efforts to adapt to the situation by attempting to provide flexible education through different methodologies, mostly online, some studies suggest that even in best-case scenarios learning was minimal, especially among young children (Engzell et al., 2021). Loss of learning was even worse for marginalised populations, who often have minimal access to technology, connectivity and adequate spaces for learning at home, and more sources of economic and psychological stress (Minz & Chhabra, 2023). Four out of five of the 104 countries who studied the effects of the pandemic on their education systems experienced learning losses (United Nations Economic and Social Council, 2023). These effects are likely to be larger on countries that do not have the capacity to undertake studies such as these.

Pending challenges

Exclusion

21. The phrasing of indicator 4.1.1, which measures the percentage of children achieving MPLs in school, does not account for those out of school. The children most likely to be out of school—such as displaced populations, refugees, the poor, and those with disabilities—are not represented in these assessments. Consequently, the indicator fails to capture the full picture of educational disparities and challenges, particularly among marginalized groups. This exclusion skews the data, likely leading to an overestimation of educational achievement, since marginalised populations are often those that struggle the most to achieve learning outcomes.

In 2019 the EiE Data Summit Action Agenda highlighted the need for a Reference Group on EiE data to tackle the sectoral challenges in EiE data collection, storage, sharing, and use. This is co-

facilitated by the International Network for Education in Emergencies (INEE), Education Cannot Wait (ECW) and UIS. However, in addition to the challenges faced across LICs and MICs, additional challenges exist in collecting data in EIE settings (Outhred *et al.*, forthcoming), including that:

- Variance in the types of conflict and the political circumstances surrounding the conflict significantly impact on how learning assessment can be implemented.
- The provision of learning in such contexts is often provided by multiple organisations, with varying degrees of focus, quality and approaches.
- The target population for education provision in conflict and crisis affected settings is extremely diverse. Often education is being delivered to internally displaced populations, refugees moving across borders and to host communities; sometimes in camps and sometimes in formal schools. These target groups bring different social and cultural norms, different languages, different curriculum exposure and potentially different future curriculum needs. At the same time, some population groups continue to be highly mobile.

22. Another significant exclusion pertains to children with visible and invisible disabilities. The first challenge is that definitions of what constitutes a disability are unclear, and existing instruments to measure it, such as the Washington Group on Disability. Statistics are contested. Because of this, the population of students suffering from disabilities is poorly identified. In addition to this, most assessments are not designed to accommodate these students, making it difficult for them to participate and perform. Universal design strategies for assessments, along with testing accommodations and modifications, are either entirely lacking or only just beginning to develop in both national and cross-national tools, which further marginalizes these children. This oversight results in an incomplete and inaccurate measure of foundational learning, ignoring the performance of a vulnerable segment of the student population (Johnstone *et al.*, 2020).

Uptake

23. Despite the emphasis of most learning assessment programmes on their relevance for evidence-based policy making, there is often a disconnect between data collection and mechanisms to ensure that this data meaningfully influences policy and pedagogy. Participation of LICs and MICs in large-scale assessments has grown considerably over time and yet research documenting the real impact of these assessments on policy remains scarce. In some cases, countries might participate in these assessments primarily for international visibility rather than for actionable insights, leading to large amounts of available data that remain unused by policymakers and educators.

24. As was mentioned above, UIS guidance on monitoring progress towards SDG 4.1.1 has indicated that international assessments should be prioritized over national ones. While the decision has been made to ensure technical rigour and favour alignment with the MPLs, this focus often comes at the expense of practical political uptake. Evidence suggests that national and even regional assessments are more likely to influence policy because they are more closely aligned with local educational contexts and needs (Tobin *et al.*, 2016). This makes it more likely for the data to provide useful information for monitoring national policies aligned with governmental priorities. For example, both SACMEQ and ERCE have been used in its respective regions to revise national grade repetition policies (Howie, 2022; Treviño *et al.*, 2016). While ILSAs are methodologically sound, their global focus does not lend itself as easily to monitoring the immediate needs of local education systems.

25. The evidence suggests that to increase the likeliness of evidence uptake, it is useful to integrate data collection into the policy-making process. This means, for example, creating assessment agencies that are independent from changes in government, legislating about the frequency of assessment or the structure of feedback data loops, and achieving sector-wide agreements on goals

related to assessments (Kaffenberger & Spivack, 2023). This integration creates predictability and generates expectations around results, which helps all stakeholders demand accountability, and thus see a response in terms of public policy. However, these conditions are more likely in high-income countries. LICs and MICs often struggle to establish reputable independent agencies capable of conducting regular assessments, either because of issues of political will, lack of financing or lack of technical capacity (Tobin et al., 2016). Data available from international agencies can be harder to integrate, because assessment cycles and priorities are established by the agencies in charge of their implementation. The lack of involvement from governments in the design process makes them less likely to legislate based on them. Without concrete mechanisms for the coherent integration of data, it is less likely that data on learning outcomes will result in policy changes.

26. Another issue affecting uptake is the quality of dissemination (Cresswell, 2017). The outputs generated from assessments are usually dense and difficult for stakeholders to understand and utilize effectively. Because of this, results often fail to reach the stakeholders they intend to inform (Tobin et al., 2016). On the other hand, when the spread of results is successful, it might lead to unintended consequences. Media coverage of learning outcomes can be hugely influential, swaying public opinion and prompting policy responses, but the difficulty in correctly interpreting results often leads to media outlets misrepresenting reports, leading to reactions that do not align with the actual findings. For example, the focus on reporting how countries rank in international assessments such as PISA regularly leads to strong reactions in public opinion and statements from politicians pressing for reform that is rarely aligned with the nuanced conclusions of the full reports (Grey & Morris, 2018). This misalignment between the intended message of the assessments and public perception can hinder the effective use of data in shaping education policy.

27. These issues also point to the political nature of learning assessments. Public opinions on what and if children are learning have political ramifications for government. The formative evaluation of the South-East Asian Primary Learning Metrics regional assessment programme found ministries of education need greater support in packaging and messaging the results of the assessment. The report states that while ministers of education often adapt and use a common language to communicate with citizens regarding examination results, they struggle to form a common and understandable language around SEA-PLM results, and link results to the intended actions of the ministry (Outhred, Turner, & Mba, 2021).

28. Despite these challenges, there is evidence of assessments having impact on policy, albeit often through top-down changes. The areas most influenced by assessment results tend to be curricula and monitoring, learning, and evaluation (MLE) systems. For example, assessments such as PISA and ICCS have prompted several countries to modernise their curricula and integrate skills aligned to global citizenship goals, such as critical thinking, problem-solving skills and civic attitudes (Tobin et al., 2016). However, changes in more critical areas such as financing, pedagogy, or teacher training are much rarer, indicating a limited scope of influence for these assessments on broader educational reform.

Sustainability

29. The sustainability of current assessments is another pressing concern. Many of the assessments deemed valid for reporting are both expensive and complex to implement. These costs and complexities are particularly burdensome for fragile states and regions experiencing educational emergencies. While some assessments include capacity-building activities, these efforts often fall short. They tend to aim for an unrealistic level of technical expertise within small national teams that suffer from high turnover and lack of institutional support (Görgens & Kusek, 2010).

30. It is often forgotten in the international development sector that regular national standardised assessments conducted in high income countries are predominantly implemented by specialised private/non-governmental organisations, rather than by governments themselves (though the secondary analysis of the results can be conducted by governments in some cases). A more sustainable approach might be to build the capacity of national governments to draft effective terms of reference and procure and manage experts according to their specific needs and resources, and focus their efforts of data analysis, interpreting findings and developing recommendations and action plans. Building this capacity would help ensure that assessments are both affordable and practical, ultimately leading to more reliable and actionable data. By tailoring support to the unique challenges of different regions, particularly those in crisis, we can create a more equitable and sustainable system for assessing foundational learning.

Annex 8 Potential Assessment Criteria of Progress

Indicative assessment criteria

- D4FL funding is sufficient to support additional countries and produce the outputs
- Learning Data Compact and other coordination mechanisms are effective
- More national governments are open to be supported to link assessment internationally
- D4FL supported interventions on the instructional core of foundational learning, trigger demand for learning assessments and vice versa
- Evidence of targeted support to data generation influencing how data is used by countries

Global learning data gaps filled, including in emergency contexts

- The acceptance by the UIS of D4FL-supported countries' data against SDG4.1.1a and b, as supported by ACER UK and the World Bank
- Acknowledgement in global forums that the PAL Network's ICAN/ICARe and other non-AMPL assessments deliver globally comparable data
- More country authorities see global reporting of learning data as important.
- Global public goods developed are used by countries not supported by D4FL
- More countries in emergencies are included in the D4FL-supported group of countries
- Children with disabilities are included in a balance of D4FL supported data generation exercises for global reporting

More disaggregated learning data available that can be used to design and monitor foundational learning interventions

- Actors (education authorities, parliament, media, teachers) in a balance of PAL Network D4FL-supported countries accept and use the PAL data
- More D4FL interventions to generate data start by engaging important country actors, including teachers, on needs from learning data
- Government and/or civil society in countries supported by D4FL make resources available or seek resources to repeat assessments regularly
- Countries identifying assessment needs and developing country protocols for conducting a suite of quality learning assessments
- Concrete country actions toward institutionalised demand for disaggregated learning data (commitment to integrate in policies/legal frameworks, integration into EMIS, routine public reporting)
- Generated data is disaggregated (in ways that are useful for the country) and already include or have plans to include children with disabilities

Increased capability in priority countries to use learning data to design and monitor foundational learning interventions.

- Teachers are aware of and interested in using learning data for classroom instruction
- Evidence of non-state actors and communities being interested in using learning data for improved instruction
- Different levels of government (national, sub-national and schools/teachers) progressing toward the use of assessment data in evaluation processes and reforms to improve foundational learning
- Countries have access to resources to implement and are interested in implementing evidenced-based reforms
- Governments feel and respond to increased pressure for action on foundational learning.
- Countries learn from one another on conducting assessments and using learning data to improve instruction
- Foundational learning remains on the global agenda as a key global education target

Annex 9 Assessing D4FL against the Evaluation Assessment Criteria

Question and sub-question	Judgement criteria	Rating	Notes
Relevance			
1. How relevant are the programme objectives to the needs of countries and global community? Objectives include the intermediate outcomes, outcomes and impact set out in the ToC			
1.1 To what extent are the programme objectives relevant to the needs of LICs and MICs?	Country governments, teachers, community leaders and other country stakeholders are interested in and working with partners on activities in countries	Green/Amber	A lot of in-country work has not begun yet, but where work has begun, partnerships are going well. Demand-led model means further demand remains to be seen, but countries have approached for support (including via UIS)
	Countries lack learning data to inform policy and practice	Green	Many countries do not have established measure for foundational learning. Case countries highlighted need for learning data to inform policy and practice.
	Countries understand the implications of and want to generate comparable and disaggregated learning data	Amber	Partners are aware of the challenges that may arise when data shows low learning levels. To date, countries seem to be on board but evidence from other initiatives points to risks and more work may be done to support use of data and encourage openness.
1.2 To what extent has support provided to date been relevant and aligned to countries' needs?	Countries have found the support to be relevant and aligned to their needs	Green	Where work has begun, government stakeholders are happy with the support and satisfied that it will meet their needs.
1.3 To what extent are the programme objectives relevant to the needs of the global community?	There are global learning data gaps	Green	Most LICs/MICs are not currently reporting internationally.
	There is a global commitment to filling these gaps in order to monitor education quality and learning outcomes	Amber	While SDGs are not legally binding, governments are expected to monitor progress, and currently global stakeholders are committed to globally comparable data as a means to monitoring education quality and outcomes. However, there are uncertainties about the requirements for internationally reporting and whether standards are reachable and a sense that momentum might be waning.
	Global learning assessment stakeholders have engaged with the D4FL programme and are interested in aligning activities and/or making use of outputs	Amber/Not known	Evidence from evaluation not complete. The evaluation spoke to Gates Foundation who are aware and interested. But not to other stakeholders.
2. How relevant is the programme's design to achieving its objectives, including on equity?			
2.1 To what extent are planned activities designed to leverage opportunities and address barriers to international reporting of comparable and disaggregated learning data?	Planned activities are appropriate and contextually relevant to participating countries	Green/Amber	To date, there are no indications that planned activities will not be appropriate and contextually relevant, and work that has started is deemed to be so. But much of the work is yet to begin. Amber for limited contextualisation of PAL Network

Question and sub-question	Judgement criteria	Rating	Notes
	Planned activities address real opportunities and barriers to international reporting in LICs and MICs	Green/Amber	Partners are responding to opportunities and planning activities to address barriers to international reporting. But it is not yet known whether the data generated will meet the UIS criteria.
2.2 To what extent are planned activities designed to leverage opportunities and address barriers to international reporting of learning data in emergency contexts?	Planned activities are appropriate and contextually relevant to emergency contexts	Amber	Only instance is Somalia. ALMA will do a AMPL in Somalia, that is contextually relevant But this is the only context where it was tested
	Planned activities address real opportunities and barriers to international reporting in emergency contexts	Red	To date, there has not been an intentional focus on the specific challenges around international reporting in emergency contexts.
2.3 To what extent are planned activities designed to leverage opportunities and address barriers to raising the quality of national assessment systems and learning data?	Planned activities are appropriate and contextually relevant to participating countries	Green Too early to say	To date, there are no indications that planned activities will not be appropriate and contextually relevant, and work that has started is deemed to be so. But much of the work is yet to begin.
	Planned activities address real opportunities and barriers to all aspects of quality including utility, credibility, integrity, validity, reliability and fairness	Green	Partners working on strengthening national assessments are responding to opportunities and intending to develop/strengthen assessments to be robust and fit for purpose.
2.4 To what extent are planned activities designed to leverage opportunities and address barriers to governments and educators using learning data to improve foundational learning?	Planned activities are appropriate and contextually relevant to participating countries	Green	To date, there are no indications that planned activities will not be appropriate and contextually relevant, and work that has started is deemed to be so. But much of the work is yet to begin.
	Planned activities address real opportunities and barriers to governments and educators using learning data to improve foundational learning, including for marginalised or disadvantaged groups, especially girls	Red	To date, focus has been on generating data that is usable. Little work is planned to address other barriers to governments and educators using data, including capability. Meanwhile, Accelerator strand does not have a consistent focus on using learning data to strengthen provision.
3. Are current processes likely to ensure countries' needs are identified and met?			
3.1 To what extent are processes for deciding who should support whom, and how, likely to ensure countries' needs are met?	FCDO and implementing partners have appropriate processes in place to decide who should support whom, and how, which will ensure countries are met	Amber	Partners have been selected for their appropriate capabilities, but this <i>ad hoc</i> approach may limit the programme's ability to fully address systemic gaps. The demand-based approach to allocating support may also limit broader systemic impact. Where FCDO and WB have a relationship, there is a risk of duplication. There is not yet a process in place to determine which countries to support if demand outstrips supply (for ACER UK support).
	To date, support is being provided by partners with the	Green/Amber	To date, stakeholders are satisfied that partners have the appropriate capacities.

Question and sub-question	Judgement criteria	Rating	Notes
	appropriate capacities to meet countries' needs		There may be gaps around supporting in-country decision-making, generating data in EiE contexts, and working on teachers' use of assessment.
3.2 To what extent are partners' processes adequately identifying and ensuring that TA meets countries' needs?	Implementing partners have appropriate processes in place to assess countries' needs	Green Too early to say	Partners' processes for identifying and ensuring that TA meets countries' needs demonstrate promising early results, with evidence of tailored scoping, contextualization, and strong partnerships. World Bank could not be fully evaluated as in-country work has not started
	Implementing partners are tailoring support provided/agreed to date to countries' needs, priorities and capacities	Green Too early to say	ACER UK is green World Bank too early to say as programmes have not started
Coherence			
4. To what extent is D4FL a coherent programme across strands and partners?			
4.1 To what extent are D4FL partners working towards common objectives?	Implementing partners' theories of change align with the programmatic ToC	Green	Align with (explicated) theory of change – explications include making Accelerator fit
	FCDO, implementing partners and countries have a shared understanding of what quality learning data means	Green/Too early to say	Alignment between FCDO and partners, countries remains to be seen.
4.2 To what extent are activities through the <i>Generate</i> strand likely to be able to meet the needs of both the <i>Track</i> and <i>Use</i> strands?	Partners delivering <i>Generate</i> strands have identified and prioritised needs of both global (<i>Track</i>) and national (<i>Use</i>) data users	Green/Amber Too early to say	They have, but unclear if sufficiently and if they will end up being compatible
	<i>Track</i> partners anticipate data generated through <i>Generate</i> activities being useful and usable	Green	
	National stakeholders anticipate data generated through <i>Generate</i> activities being useful and usable	Green	Limited but yes so far
5. To what extent are D4FL activities compatible with in-country education and assessment systems, structures and policies? (i.e. external coherence at country level)			
5.1 To what extent are in-country programme activities aligned with national education structures and policies?	D4FL activities generate evidence to inform national education structures and policies	Green	(plans)
	The logistical features (e.g. timing, mode) of assessment activities established through D4FL align with national education structures	Green	(plans)
5.2 To what extent are the objectives and approaches of	D4FL activities support the right functions within countries' assessment systems to produce	Green / Too early to say	For ACER UK, where activities have started, possible to say. For others no

Question and sub-question	Judgement criteria	Rating	Notes
D4FL activities coherent within countries' assessment systems (including assessment for learning, examinations, citizen-led assessments, sample-based standardised assessments)?	learning data that meets the needs of the <i>Track</i> and <i>Use</i> strands		engagement with government on their programmes as it is too early
	D4FL is not duplicating efforts	Green/Amber	We have examples of D4FL partners avoiding duplication, but also examples of partners not aware of other activities
6. To what extent are D4FL and other FCDO projects complementing one another? (i.e. internal FCDO coherence at country and global levels)			
6.1 To what extent are D4FL and other FCDO projects actively engaging and complementing one another in-country?	D4FL and other projects are learning lessons from one another	Red	Not happening yet
	D4FL and other projects are aware of one another and how their activities fit together	Amber	One example of D4FL and GPE in country. But elsewhere not aware or no clear understanding of how they fit together
	D4FL is providing guidance to FCDO education support programmes on generating and using learning outcome data	Red	Not happening yet
6.2 To what extent are D4FL and other FCDO projects actively engaging and complementing one another at a global level?	D4FL and other projects are learning lessons from one another	Red	Not happening yet
	D4FL and other projects are aware of one another and how their programmes can fit together	Amber	Some awareness at global level. But not clear understanding of how activities fit together
Efficiency			
7. Will the programme set-up likely facilitate timely and efficient use of resources?			
7.1 To what extent are current processes (contracting mechanisms, contractual arrangements and partners internal processes) likely to facilitate <u>timely</u> use of resources?	Processes are likely to facilitate timely responses to requests for support (within planned timeframes where applicable, and within timeframes that align with priorities for support and are acceptable to countries)	Green Too early to say	Partners have different timelines, but are on track for their timelines. Where partners respond directly, the response is fast. Where activities have not started, the evaluation did not get an opportunity to explore what would be acceptable to countries.
	Processes are likely to facilitate timely delivery of support (within timeframes that align with priorities for support where realistic, and at a pace that is acceptable countries and does not risk quality and relevance)	Green Too early to say	Where activities have started they are on track. But, sizable parts of the programme have not yet started activities on the ground.
7.2 To what extent are current processes (contracting mechanisms, contractual arrangements and partners internal processes) likely to facilitate <u>efficient</u> use of resources?	Processes are in place to ensure work by partners is well managed	Green	Processes are in place and working, both within partners and between partners and DFID. No evidence of material issues.
	Processes in place are likely to minimise waste or duplication of efforts	Green/amber	In principle, direct sourcing may mean less opportunity to achieve value for money, but with each partner, sufficient opportunity in implementation to ensure efficiency

Question and sub-question	Judgement criteria	Rating	Notes
			The lack of systematic coordination between partners raises a risk of duplication.
	Processes are in place to ensure that resources are targeted at meeting programme outcomes and country priorities	Green	There are contracts and results frameworks in place with all contracted partners. For the most part partners are working towards their contracted outputs.
7.3 To what extent and how are partners prioritising efficient use of resources?	Partner organisations have prioritised efficiency in their planning	Green	For each partner there are efficiency-positive factors
	Partner organisations have put appropriate processes and protocols in place to ensure that resources are used efficiently (as defined above)	Green	From the data collected, no evidence to suggest resources are being wasted. All three partners have robust management.
8. To what extent are partners responding in a timely way to requests for support?			
8.1 To what extent are partners responding in a timely way to initial requests for support?	Partners are responding to initial requests within planned timeframes (where applicable)	Green	Partners are responding within timeframes.
	Partners are responding to requests within timeframes that align with countries' priorities for the support	Green/amber	Some partners yes, but others' internal processes determine the time frames for response
	Countries are satisfied with the timeliness of responses	Green Too early to say	Where activities have started, countries are satisfied. Out of 4 countries we were able to check with 2 sets of country authorities.
	Countries remain engaged with the process	Green Too early to say for all partners	Where activities have started and we have engaged country authorities, they are definitely engaged. For other countries, we have understood from global teams that countries are engaged.
8.2 To what extent are partners working with countries to agree support packages in a timely way?	Partners agree support packages within planned timeframes (where applicable)	Green Too early to say for all partners	Where activities have started, support packages are agreed with government swiftly.
	Partners are agreeing support packages within timeframes that align with countries' priorities for the support	Green Too early to say for all partners	Where activities have started, and we've engaged governments, the timeframes are aligned with their priorities.
	Countries are satisfied with the pace of progress	Green Too early to say for all partners	Where activities have started, and we've engaged governments, countries are satisfied. For other, global partners have indicated satisfaction.
	Countries remain engaged with the process	Green Too early to say for all partners	
	Quality and relevance of support packages agreed are not impaired by a desire to move quickly	Green Too early to say for all partners	Where activities have started and we have engaged country authorities, they are definitely engaged. For other countries, we have understood from global teams that countries are engaged
9. To what extent have D4FL resources been used efficiently to date, and its intended timeline kept?			
9.1 To what extent have partners used D4FL resources efficiently to date?	Partners' work has been well project managed to result in quality outputs being delivered within expected level of effort	Green	There is good evidence of good project management, and no evidence of shortfalls

Question and sub-question	Judgement criteria	Rating	Notes
	Partners' work has involved minimal waste and duplication of efforts	Green/amber	There is some duplication, as noted under coherence
	Partners have targeted resources at meeting programme outcomes and country priorities	Green/amber	Most partners have. Some partners have prioritised countries, and while the impact will be met, the programme outcomes as stated are less certain to be supported.
9.2 To what extent has FCDO's allocation and management of resources across strands and partners been efficient?	Minimal waste and duplication of efforts across strands and partners	Green/amber	While resources are targeted at outcomes, there is a risk the D4FL is doing too much, and will not achieve significant enough outcomes on any one objective. Not material – unless improvements on design, coherence and addressing barriers to progress are not realised.
	Resources have been used as planned	Green	Resources are used as planned
9.3 Is the programme being delivered according to the intended timeline?	The programme timeline is being kept. If the programme timeline is not kept, the deviations are merited and likely to support better achievement of outcomes.	Green	The programme timeline is being kept. The small deviations are merited by processes to target resources better / processes to get countries on board.
CRITERION: EFFECTIVENESS			
10. To what extent are D4FL activities aligned with the programme ToC?			
10.1 To what extent are plans and activities to date aligned with intended outcomes, including equity outcomes?	Track strand plans and activities to date are focused on producing learning data that: <ul style="list-style-type: none"> · is comparable · is disaggregated, including by gender and disability · includes emergency contexts 	Amber	There is good alignment between the ToC outcomes and some activities. Other outcomes are under-supported by activities, and some activities are not well represented on the ToC.
10.2 To what extent are planned activities designed to leverage opportunities and address barriers to international reporting of comparable and disaggregated learning data?	Generate and Use strand plans and activities to date are focused on supporting LICs and MICs to generate data that is <ul style="list-style-type: none"> · robust · disaggregated, including by gender and disability · useful Generate and Use strand plans and activities to date are focused on supporting governments and teachers to use learning data to improve foundational learning		
10.3 To what extent are plans and activities under the generate and use strands aligned with the guiding principles?	Plans and activities to date are focused on providing support that is nationally and sub-nationally owned; and answers problems faced by countries, and effectively responds to their own demands and targets	Green/Amber	Partner interviews and documentation demonstrate commitment to government ownership for the most part. Interviews with country authorities have confirmed. But some processes appear to be driven by partner needs, and this criterion will only be met if the partner needs is fully based on country needs and problems.
	Plans and activities to date are focused on providing support that includes teachers and schools leaders and supports them to effectively do their job	Red	There are very few activities that set out to include teachers and school systems, except for Accelerator that are aimed at teacher capacity building (on instruction) and professional development, alongside

Question and sub-question	Judgement criteria	Rating	Notes
			activities on curriculum and teaching materials.
	Plans and activities to date are focused on providing support that is cost-effective and sustainable (i.e. not too burdensome for governments, schools and teachers), but allows for rigour, quality and comparability	Green/amber Too early to say	Good evidence of cost-effectiveness (because the D4FL leverages other funding) but sustainability is not yet proven.
	Plans and activities to date are focused on providing support that is effective at building state capability (not reliant on external or parallel processes)	Green	All partners' plans include building state capability, even the PAL Network, where the aim is to build capability through accountability and engagement.
11. What has the programme achieved to date?			
11.1 Are partners making expected progress in delivering activities and outputs?	Implementing partners are making expected progress in delivering activities and outputs	Green/amber	If outputs only, programme is on track. But, from an outcome perspective, there is only a few 'use' activities in evidence.
	Outputs are being achieved within expected level of effort	Green/amber Too early to say	There is some evidence of higher cost than expected in the citizen-based data generation strand. Too early to say for activities that have not yet started.
11.2 What results have been achieved to date regarding filling global learning data gaps?	More LICs and MICs are reporting internationally using comparable and disaggregated data on learning, including in emergency settings	Green/amber Too early to say	There will be incremental progress with regards to this objective. At present it is highly likely that at least some countries will report that have not reported before. But acceptance by the UIS is not tested for any one D4FL initiative. Also, if about proof of concept, yes, but for systematically more countries reporting, even countries not helped by D4FL, unsure.
11.3 What results have been achieved to date regarding the generation of learning data in-country?	Rigorous, regular primary learning assessments are being implemented	Green	Where activities have started, there is evidence of assessments being improved, even if UIS does not accept data for global reporting
	Learning assessment is being embedded in national policies and reforms	Amber	There are limited cases. The creation of Assessment Frameworks with D4FL support is one example, and efforts known from ALMA to create information tools that will demand data another.
	Simple foundational learning assessment is being used in classrooms and community interventions	Red	There is no evidence of planned or ongoing activities.
11.4 What results have been achieved to date regarding the use of learning	Learning data is being used by government to implement and monitor interventions	Red	There is no evidence of planned or ongoing activities to support this, apart from in 3 of the 12 Accelerator interventions, where learning data and

Question and sub-question	Judgement criteria	Rating	Notes
assessments to strengthen curriculum and instruction?			evaluation are explicitly part of the activities.
	Learning data is being used by teachers and other educators to strengthen practice	Red	See above.
12. What are the likely barriers and facilitators to achieving results?			
12.1 What are the likely barriers and facilitators to achieving results regarding filling global learning data gaps?	Assessment will be linked to emerging findings against earlier EQs and will include assessment of the validity of the key ToC assumptions	Green	The assumptions are valid, in that they describe key barriers (if negative) and facilitators (if positive) to achieve programme outcomes.
12.2 What are the likely barriers and facilitators to achieving results regarding the generation of learning data in-country?	Assessment will be linked to emerging findings against earlier EQs and will include assessment of the validity of the key ToC assumptions	Green	See above
12.3 What are the likely barriers and facilitators to achieving results regarding the use of learning assessments to strengthen curriculum and instruction, including for marginalised groups?	Assessment will be linked to emerging findings against earlier EQs and will include assessment of the validity of the key ToC assumptions	Red	The ToC does not articulate clearly what the assumptions and pathways are to results on use of learning data.
LOOKING AHEAD AT IMPACT AND SUSTAINABILITY			
13. What are the likely pathways and timelines for outcomes and desired impact?			
13.1 What are the likely timelines for the outcomes and desired impacts articulated in the D4FL ToC?	FCDO and partners' expected timelines for outcomes and impacts are realistic	Amber	Incremental achievements yes, but not systematic achievement of outcomes and impact within programme timeframe.
13.2 What are the expected pathways to intended outcomes, and therefore what early indicators of promise should the summative evaluation aim to capture?	Pathways in the final explicated Theory of Change for testing in the summative evaluation, are clear and credible Summative evaluation can provide evidence of expected pathways and early indicators of promise within the duration of the evaluation	Amber	The ToC has gaps, and some pathways are not clear (e.g. on use)
14. What are the avenues for D4FL to have sustainable impact?			
14.1 How do FCDO and partners conceptualise sustainability for D4FL?	Clear and credible conceptualisation of how outcomes achieved through the D4FL, will be sustained Conceptualisations of sustainable impact are clear and credible Conceptualisations of sustainable impact align with countries' and global sector's needs	Amber	Conceptualisation of sustainability and impact is there but not consistently across partners

Question and sub-question	Judgement criteria	Rating	Notes
	FCDO and partners' conceptualisations are aligned		
14.2 To what extent and how has sustainability been integrated into programme design?	Sustainability has been appropriately integrated into programme design	Amber	Sustainability can be better integrated into programme design.
14.3 What other opportunities are there for sustainable impact?	n/a		
15. What are the avenues for FCDO to influence the wider sector?			
15.1 How does FCDO hope to influence the wider sector?	Aspirations for influence clear and credible Aspirations for influence align with global sector's needs FCDO and partners' conceptualisations are aligned	Amber/Red	Aspirations are perhaps more opportunistic, rather than clear, though opportunities for influencing the wider sector proactively are there.
15.2 How can learning and outputs from D4FL benefit the sector?	n/a		

Annex 10 Explicated Recommendations

Table 14 Recommendations

Recommendation	Rationale	Responsibility and timing
Recommendation 1: Adjust the Theory of Change to reflect current programme design and intended mechanisms of change		
<p>Adjust the Theory of Change to reflect the currently intended inputs, outputs, intermediate outcomes, outcomes and impact of the programme, and how they are expected to lead to one another.</p>	<p>An accurate ToC will enable the summative evaluation to accurately evaluate the programme against its intended outputs and outcomes. In addition, the ToC can serve as a programming tool, helping partners to understand how the different strands of work fit together, and enabling learning conversations about how well the different strands of work are progressing and how they might better join up.</p> <p>In Annex 4 we have proposed an explicated ToC that better articulates the current programme design, including:</p> <ul style="list-style-type: none"> a) how activities funded through the Accelerator initiative are intended to contribute to the ultimate outcome of improved foundation learning, outside of the causal pathways around strengthening the generation, reporting and use of learning data (see Chapter 4) b) the pathways through which the programme outcomes around the generation, reporting and use of learning data are expected to lead to that ultimate outcome (see Chapter 4) c) the ways in which the “proof of concept” provided by D4FL and increased pressure for action on foundational learning expected to result from the availability of learning data are expected to increase demand for learning data, thus reinforcing programme outcomes (see Chapter 4) d) mapping assumptions (including additional assumptions articulated by implementing partners) onto the different stages or levels of the ToC (inputs to outputs, outputs to outcomes, outcomes impact) <p>We propose to finalise this ToC with FCDO and use it as the basis for the summative evaluation framework.</p>	<p>FCDO, immediate</p>
Recommendation 2: Articulate the relative importance of national and global programme objectives		

<p>Consider and decide/articulate more clearly the relative priority of countries having any useful data on foundational learning vs. meeting global reporting requirements.</p>	<p>At the moment, the filling of global learning data gaps and the availability of disaggregated learning data that can be used to strengthen education provision in country are given equal weight in the Theory of Change (see Annex 4). <i>Generate</i> activities are more oriented to <i>Track</i> than to <i>Use</i>. As discussed in Chapter 4, partners are optimistic that the data they will generate or countries to generate through D4FL can meet both needs. However, learning from previous initiatives demonstrates that global and in-country needs can be conflicting. Moreover, the pathway to the ultimate outcome of improving foundational learning necessarily requires in-country action, while the causal pathway from global reporting to improved learning outcomes is more protracted and indirect (see Chapter 4, Annex 4). There is also some ambiguity as to whether filling “global learning data gaps” refers specifically to reporting through UIS mechanisms, or a broader category of data that is internationally comparable.</p> <p>A clearer articulation of this goal, and the relative importance of national use vs. global reporting, will help the summative evaluation to more accurately report on the programme’s achievements against its ambitions, as well as supporting FCDO and D4FL implementing partners to continually reflect and improve on what they are achieving and prioritise efforts.</p> <p>For example, articulating this clearly will help implementing partners to make decisions about what to prioritise in instances where global and national needs might be in conflict (e.g., how many and which learning domains to include in national assessments that countries want to use for global reporting, whether the PAL Network should focus on processes that enable the generation of globally comparable data in more countries, or data that can be more useful for in-country decision making, at a higher per-country cost).</p> <p>Articulating priorities will also have implications for the selection of countries to receive support. For example, if the primary goal is global reporting, it would make sense to select countries with relatively established assessment systems and strong in-country capacities, where there is potential for TA to “tip” countries over the threshold of UIS requirements. If the primary goal is to strengthen the availability of data for use in-country and capacity to use that data, there is an argument to support countries with less already in place.</p>	<p>FCDO, immediate</p>
--	---	------------------------

	<p>Finally, articulating this order of priorities will help FCDO to make decisions about which, of any potential, new strands of work to fund through D4FL – focusing on helping countries to have any (comparable) data on foundational learning vs. helping countries to report against SDG4 through global reporting mechanisms.</p>	
<p>Recommendation 3: Remove classroom use of assessment as an intended activity strand and shift to a focus on teachers as a key stakeholder group for D4FL <i>Generate</i> and <i>Use</i> activities</p>		
<p>Reorientate thinking away from use of assessment in classroom and community interventions, towards prioritising teachers as a key stakeholder group in national assessment systems – as users of data and experts in how children learn and what they need to know and be able to do in terms of foundational learning.</p> <p>We recommend that FCDO and relevant implementing partners have a workshop in early 2025 to consider how teachers can be made more central in project efforts and what will be feasible within the time and resources available for the remainder of the project.</p>	<p>The current ToC includes “Strengthened use of simple foundational learning assessment in classroom and community interventions”. However, there are currently no planned activities to bring about this output (see Chapter 4, Annex 4). This would be a substantial new strand of work and would likely require bringing in new partners to ensure the right capabilities would be included. It is unlikely that the time and budget remaining for the D4FL would be make substantial progress towards results in this area. We therefore propose that this is out of scope of the programme and should be removed from the ToC. The wording of the final input in the ToC (currently “<i>Support to bilateral and CMPs to ensure quality assessment for learning applied in interventions to support instruction</i>”) should also be reconsidered.</p> <p>It is possible that the Accelerator initiative will include interventions focused on classroom use of formative assessment, as one aspect of pedagogy that can be strengthened to improve foundational learning outcomes. However, this is not currently a focus.</p> <p>Teachers are an important stakeholder group. We recommend that efforts and resources are put towards ensuring that partners’ in-country work on planning for learning measurement and using learning data includes teachers and/or supports national governments to consult with teachers on the data needs in their context, to disseminate findings to teachers and to include teachers in decision-making about what “results” should mean for action. This could be included as an assumption in the ToC.</p>	<p>FCDO, immediate Partners, 2025-2027 in designing and implementing activities</p>
<p>Recommendation 4: Increase the focus on the inclusion of EiE settings as a core programme objective</p>		
<p>Implementing partners to increase intentional focus on EiE. FCDO and partners to work together to develop a typology for assessing EiE contexts, to</p>	<p>The current programmatic ToC forefronts emergency contexts in intermediate outcomes and outcomes. But this was not identified as a particular focus for any of the implementing partners (though some partners are working in EiE contexts, see Chapter 4), and so there is a disconnect between the ToC on paper and working theories of change and intended outcomes. Currently, the programme would be deemed to have not</p>	<p>FCDO, immediate; Partners, 2025-2027 in designing and implementing activities</p>

<p>identify what will be feasible, which in-country stakeholders to work with, and what additional expertise might be needed. FCDO to link implementing partners to resources available on conducting assessment in emergencies and to work with global partners on improving the architecture and evidence available on learning data generation in emergency and crisis contexts.</p>	<p>met its intended goals if no emergency contexts were included in global reporting, for example.</p> <p>Another fund (Education Cannot Wait) holds the remit for education and learning measurement in emergencies and protracted crises. Moreover, the programme activities and delivery models are likely to be suitable for some emergency contexts but not others. For instance, adjusting national assessments to meet UIS requirements will enable global reporting to include refugee and displaced learners in country contexts where government are responsible for the education of these groups, but not in contexts where education for refugee and displaced populations is provided by non-state actors.</p> <p>We therefore propose that that implementing partners, supported by the FCDO, should increase their intentional focus on EiE, tailored for the context in which they are working and their capabilities, and drawing on existing resources on conducting assessments in emergencies.</p>	
<p>Recommendation 5: Improved joint thinking and cooperation vertically between FCDO and partners and horizontally between partners at the global and country levels</p>		
<p>The FCDO should think strategically about how it can influence all partners to pull together for programme outcomes. The FCDO and partners should develop shared principles on coordination/cooperation that can guide expectations and responsibilities going forward.¹⁸</p> <p>The FCDO should facilitate more joint thinking and cooperation between partner at the global level, on how to achieve the programme objectives jointly, and address shared challenges.</p> <p>Partners should invest in cooperation with the frame of the D4FL, for cross-</p>	<p>Closer cooperation between partners (including all <i>Track, Generate</i> and <i>Use</i> partners) will contribute to the programme’s strands working together better. The recommendation is based on the conclusion, and supporting findings, that while each strand of the programme will contribute to programme results (impact and outcomes) individually, the strands are not working together well enough to optimize how the D4FL investments, will achieve the outcome results.</p> <p>The recommendation is aimed at shifting the incremental contribution that each partner will make towards the programme objectives, to a more systematic contribution that will significantly move the global landscape and country landscapes on tracking, generating and using learning data for improved outcomes forward.</p> <p>The recommendation is also aimed at addressing the findings on actual and perceived duplication of effort between partners at country level, and building on the positive examples of how coordination and cooperation can avoid duplication and bring the different capacities of partners together in the interest of countries making progress.</p>	<p>FCDO</p> <p>All implementing partners</p> <p>2025-2027</p>

¹⁸ This recommendation draws on the Learning and Reflection Partner workshop discussions.

learning and coordination between strands.

This can be achieved by:

- Regular get-togethers – i.e. twice a year -- between partners' global teams to reflect on programme challenges and learning, and to discuss how the programme can achieve shared objectives better through partners' individual actions.
- Facilitation of dialogue between partners on specific issues where individual partners may have expertise: this may be done by partners linking directly in a de facto community of practice, and the FCDO making linkages between partners where it spots opportunities.
- Think about how the work of the D4FL in the different strands can be showcased to encourage adoption and adaptation in other countries.
- Facilitate peer to peer support between partners counterparts across counties, to address challenges experienced by country actors

The FCDO should facilitate better coordination between partners at country level. This can be done by:

- Facilitating information flows between partners on who works where and on

The recommendation is aimed at improving the extent to which partners address the barriers and leverage the facilitators for progress on learning assessment, the use of data and improved learning outcomes. Different partners have different capabilities (linked to their capacities, relationships and where they sit in countries' political and institutional landscapes) to address these barriers, and coordination/cooperation can help these capabilities working together for better learning outcomes. Secondly, there is value in facilitating lessons learning between countries. Different partners may be working with similar counterparts between countries, and better coordination and cooperation can help leverage progress in country a, to address challenges in country b by connecting people and sharing of experiences.

Partners face constraints on the extent to which coordination and cooperation is possible. For example, there may sensitivities about differences between World Bank and civil society roles at country level and how they respectively relate to government and other actors, like the legislature and the media. Global teams may also have limitations on the extent to which they can enlist country-level leads/partners to coordinate with other D4FL partners. It is therefore also important to recognize these constraints, and be clear on how coordination and cooperation will work, and on role and responsibilities.

<p>what, and with what progress to enable country-level links</p> <ul style="list-style-type: none"> Partners communicating with each other on country level challenges, following a 'step up step back' approach to ensure the right partners are stepping up/stepping back according to comparative advantage 		
<p>Recommendation 6: Partners should actively strengthen how they work politically to shift political and institutional barriers to regular assessment and use of data, and the FCDO and partners should invest in joint learning on how best to do so.</p>		
<p>The FCDO and partners should invest in learning on how partners have successfully worked politically, to address existing and potential political and institutional barriers to imbedding learning assessments and use of data.</p> <p>Partners should consider how they can, at country level, work beyond their technical activities to shift how assessments and learning data are seen and used. Examples are:</p> <ul style="list-style-type: none"> Thinking how country leaders and institutions can frame the purpose of learning assessments, and tracking performance, to support sustainable assessment and use of data. Consider what support is needed to communicate assessment results. Consider how to work politically within the constraints of the support, and generate demand from the bottom up. (E.g. can like-minded government 	<p>Ensuring assessments are sustainable requires addressing political and institutional barriers to countries regularly undertaking assessments and using the resulting data to design and implement reforms in the instruction core of foundational learning.</p> <p>These barriers include the hesitancy to undertake assessments again when the previous results reflected negatively on learning performance; perceptions that global assessments are not country-relevant; perceptions that assessments are only about accountability; and change in political and technical leadership with in-coming leaders potentially not supporting regular assessments or having needed knowledge.</p> <p>Moreover, the evaluation has found that whether countries use learning data, is determined by thinking about use for education policy from the start.</p> <p>This recommendation is aimed at leveraging existing D4FL partnerships and contracted activities, to strengthen the links between <i>Track</i> – where a lot of investment is -- and <i>Generate</i> and <i>Use</i>, the engine room of learning improvement.</p>	<p>FCDO Implementing partners 2025-2027</p>

<p>actors work with committees of the legislature to emphasise why tracking learning is important for demand from that level; are there media or civil society actors in scope; what development partners can help support demand, etc.).</p>		
---	--	--

OFFICIAL

129

OFFICIAL

Annex 1 Bibliography

Full reference
Addey, C. (2024). Intensions, tensions and resistance: International organizations' struggles in monitoring and measuring SDG4. <i>International Journal of Educational Development</i> , 104, 102969-. https://doi.org/10.1016/j.ijedudev.2023.102969
Addey, Sellar, Steiner-Khamsi, Lingard, & Verger (2017) The rise of international large-scale assessments and rationales for participation
Cresswell, J. C. (2017). Dissemination and Reporting. In P. Lietz, J. C. Cresswell, K. F. Rust, & R. J. Adams (Eds.), <i>Implementation of Large-Scale Education Assessments</i> (1st ed., pp. 424–435). Wiley. https://doi.org/10.1002/9781118762462.ch17
Crouch, L., & Vinjevold, P. (2006) South Africa: Access before quality, and what should we do now?
Education Endowment Foundation (2022) Implementation and process evaluation guidance for EEF evaluations. London: EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf?v=1660741017 .
Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. <i>Proceedings of the National Academy of Sciences</i> , 118(17), e2022376118. https://doi.org/10.1073/pnas.2022376118
FCDO (2022) Business Case: Data for Foundational Learning Programme
FCDO (2024) Data for Foundational Learning Programme: 2024 Annual Report
GAML (2018) Global Alliance to Monitor Learning: 2018 Progress report
Görgens, M. G., & Kusek, J. Z. (2010) Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit. World Bank Publications.
Grey, S., & Morris, P. (2018) PISA: Multiple 'truths' and mediated global governance. <i>Comparative Education</i> , 54(2), 109–131. https://doi.org/10.1080/03050068.2018.1425243
Gustaffson, M. (2022) Trends in learning proficiency in the last twenty years: How close are we to reliable regional and global SDG 4.1.1 trend statistics? [Report prepared for the UNESCO Institute for Statistics]. UIS. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2023/02/SDG-4.1.1-historical-trends-FINAL.pdf
Howie, S. J. (2022) Regional Studies in Non-Western Countries, and the Case of SACMEQ. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), <i>International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings</i> (pp. 419–468). Springer International Publishing. https://doi.org/10.1007/978-3-030-88178-8_18
Johnstone, C. J., Schuelka, M. J., & Swadek, G. (2020) Quality Education for All? The Promises and Limitations of the SDG Framework for Inclusive Education and Students with Disabilities. Brill. https://doi.org/10.1163/9789004430365_004
Kaffenberger, M., & Spivack, M. (2023) System coherence for learning: Applications of the RISE education systems framework. https://www.elgaronline.com/edcollchap-0a/book/9781802205930/book-part-9781802205930-18.xml
Lockheed & Wagemaker (2013) International Large-Scale Assessments: Thermometers, Whips or Useful Policy Tools?
Minz, N. K., & Chhabra, I. (2023) Investigating the Impact of Covid-19 On Achieving Sustainable Development Goal 4: Education for All. Zenodo. https://doi.org/10.5281/ZENODO.8399737
Mokoro Limited and Oxford MeasurEd, 2024, Final Incept Report; Evaluation of the FCDO Data for

Foundational Learning Programme, unpublished.
Montoya, G. E. M. (2024) On the way forward for SDG indicator 4.1.1a: Setting the record straight. World Education Blog. https://world-education-blog.org/2024/03/22/on-the-way-forward-for-sdg-indicator-4-1-1a-setting-the-record-straight/
Outhred, R., Sangar, I., Nicolai, M., Valenzuela, P., and Santcross, N. (2024) ECW White paper on Measuring Holistic Learning Outcomes in Conflict and Crises Affected Settings. December 2024.
Outhred, R., Turner, F., & Mba, E. (2021) Evaluation of SEA-PLM. Cambridge: Cambridge Education.
Ritchie, J. et al. (2013) Qualitative Research Practice: A Guide for Social Science Students and Researchers.
Rustad, S.A. (2024) Conflict Trends: A Global Overview, 1946–2023. PRIO Paper. Oslo: PRIO.
Sayed, Y., & Moriarty, K. (2020) SDG 4 and the 'Education Quality Turn': Prospects, Possibilities, and Problems. In A. Wulff (Ed.), Grading Goal Four (pp. 194–213). Brill. https://www.jstor.org/stable/10.1163/j.ctv2gjwt3.15
Spaull, N. and Taylor, S. (2013) The effects of rapidly expanding primary school access on effective learning: The case of Southern and Eastern Africa since 2000.
Tobin, M., Nugroho, D., & Lietz, P. (2016). Large-scale assessments of students' learning and education policy: Synthesising evidence across world regions. Research Papers in Education, 31(5), 578–594. https://doi.org/10.1080/02671522.2016.1225353
Treviño, E., Villalobos, C., Baeza, A., & Caribe, U. O. R. de E. para A. L. y el. (2016). Recomendaciones de políticas educativas en América Latina en base al TERCE. MINISTERIO DE EDUCACION. https://repositorio.minedu.gob.pe/handle/20.500.12799/4513
UNESCO (2015) Transforming our World: The 2030 Agenda for Sustainable Development Department of Economic and Social Affairs. https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981
UNESCO (2017) Unpacking Sustainable Development Goal 4: Education 2030; guide—UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000246300
UNESCO (2018) Minimum proficiency levels (MPLs): Outcomes of the consensus building meeting. https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/12/4.1.1_29_Consensus-building-meeting-package.pdf
UNESCO (2022) Monitoring of the Sustainable Development Goals using Large-scale International Assessments: A strategy for reporting SDG 4 indicators using data from cross-national assessments. https://unesdoc.unesco.org/ark:/48223/pf0000389176
UNESCO (2023) 2023 SDG 4 scorecard report on progress towards national SDG 4 benchmarks
UNGEI (2023) G7 Global Objectives Report 2023
UNICEF (2021). COVID-19 and School Closures: One year of education disruption.
UNICEF (2024) Foundational Learning Action: Findings from the Foundational Learning Action Tracker 2024
United Nations Economic and Social Council (2023). Progress towards the Sustainable Development Goals: Towards a rescue plan for people and planet (Report of the Secretary-General (Special Edition) A/78/80-E/2023/64; General Assembly Economic and Social Council)
Wagner, Wolf, & Boruch (2018) Learning at the bottom of the pyramid: science, measurement, and policy in low-income countries
World Bank (2024) ALMA Grant Application Form FY2025
World Bank FLC (unpublished.) How FLC Funded Programs Work Together: ALMA, Accelerator 2.0,

and Implementation Science.

Annex 12 Acronyms

ACER UK	ACER International United Kingdom (ACER UK)
ALMA	Accelerating Learning Measurement for Action
AMPL	Assessment for Minimum Proficiency Levels
ASER	Annual Status of Education Report
BESTA	Better Education Statistics and Global Action to Improve Learning
BMGF	Bill & Melinda Gates Foundation
D4FL	Data for Foundational Learning
ECW	Education Cannot Wait
EdGE	Education, Gender and Equality Directorate
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EiE	Education in Emergencies
EiEPC	Education in Emergencies and Protracted Crises
EQ	Evaluation Question
ET	Evaluation Team
FCDO	Foreign, Commonwealth and Development Office
FGD	Focus Group Discussion
FLC	Foundational Learning Compact
GAML	Global Alliance to Monitor Learning
GAMLA	Global Alliance for Measuring Learning Assessment
GED	Girls' Education Department
GEM-R	Global Education Monitoring Report
GPE	Global Partnership for Education
ICAN	International Common Assessment of Numeracy
ICARe	International Common Assessment of Reading
ILSA	International Large-Scale Assessments
INEE	Inter-Agency Network for Education in Emergencies
INGO	International Non-Governmental Organization
IR	Inception Report
KNEC	Kenya National Examinations Council
LEGs	Local Education Groups
LIC	Low-Income Country

LMIC	Lower Middle-Income Country
MIC	Middle-Income Country
MPL	Minimum Proficiency Level
NaCCA	National Council for Curriculum and Assessment
NFER	National Foundation for Education Research
NGOs	Non-Governmental Organisations
OECD-DAC	Organisation for Economic Cooperation and Development's Development Assistance Committee
PAL	People's Action for Learning
PASEC	Programme d'analyse des systèmes éducatifs de la Confemen
PISA	Programme for International Student
PM	Programme Manager
QASD	Quality Assurance and Standards Department
QS	Quality Support
RELI	Regional Education Learning Initiative
SACMEQ	Southern and Eastern Africa Consortium for. Monitoring Educational Quality
SDG	Sustainable Development Goal
TA	Technical Assistance
TIMSS	Trends in International Mathematics and Science Study
ToC	Theory of Change
TOR	Terms of Reference
TTL	Task Team Leader
UIS	UNESCO Institute of Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNGEI	UN Girls' Education Initiative