



Commissioned by FCDO

Evaluation of the Data for Foundational Learning (D4FL) Programme

Final Mid-Term Evaluation Report
6 March 2026

In partnership with



Authors:

Alta Fölscher, Team Leader

Lydia Marshall, Deputy Team Leader

Christine Fenning, Evaluator

Javier Pereira, Evaluator

Rachel Outhred, Principal Expert

Paulina Valenzuela, Technical Expert

Tal Shalson, Research Coordinator

Ellen Smith, Researcher

Sara Bashir Malik, 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	v
1. Introduction	1
1.1 Scope and purpose of the evaluation	1
1.2 Organisation of the evaluation report.....	2
2. The Data for Foundational Learning Programme and its Context.....	2
2.1 Context.....	2
2.2 The Data for Foundational Learning Programme	6
3. Evaluation Methodology and Limitations.....	10
3.1 Overview of the approach.....	10
3.2 Methodology.....	10
3.3 MTE Methodological Limitations.....	12
4. Evaluation findings	13
4.1 Relevance	13
4.2 Coherence	20
4.3 Effectiveness	26
4.4 Efficiency	34
4.5 Impact	41
4.6 Sustainability	43
5. Conclusions and Recommendations	48
5.1 Conclusions	48
5.2 Recommendations.....	52
Annex 1 Terms of Reference.....	56
Annex 2 Timeline of Events in the Data for Foundational Learning Context ...	76
Annex 3 Education Assessment in the Context of SDG 4.....	77
Annex 4 Original and Updated Theory of Change, Partner Theories of Change...	88
Annex 5 D4FL implementation overview	97
Annex 6 Mid-Term Evaluation Methodology	104
Annex 7 Mid-Term Evaluation Framework	121
Annex 8 Use and Influence Plan.....	132
Annex 9 Survey Results	136
Annex 10 Programme performance against the assessment criteria	148
Annex 11 Bibliography	167
Annex 12 Acronyms.....	170

List of Tables

Table 1	MPLs for primary education.....	4
Table 2	Summative evaluation criteria and questions (MTE).....	10
Table 3	Instances of D4FL support that build links from data generation to use.....	28
Table 4	Country cases where D4FL support is likely to lead to use of data	31
Table 5	Typical delays in programme implementation	39
Table 6	Budget execution by partners mid-to-late 2025, GBP.....	40
Table 7	Primary education MPL and equivalencies with cross-national assessments	80
Table 8	Availability of data to report against indicator 4.1.1 across regions	82
Table 9	D4FL Implementing Partners, Contracted Work and Work Done	97
Table 10	Summative evaluation questions (MTE).....	105
Table 11	Evaluation matrix for the MTE	108
Table 12	Country cases	112
Table 13	MTE respondent selection for data collection	114
Table 14	Progress on use and influence plan.....	133
Table 15	Survey respondents by affiliation	137
Table 16	Perceived government priority per education stage	138
Table 17	Importance of improving access, quality and data on foundational learning.....	139
Table 18	Reasons why improving data was not considered important.....	140
Table 19	Ways to improve quality of foundational learning	141
Table 20	Barriers to reporting	142
Table 21	Stakeholder views on ACER UK.....	143
Table 22	Stakeholder views on PAL Network	143
Table 23	Stakeholder views on WB-ALMA	145
Table 24	Stakeholder views on WB-Accelerator 2.0.....	145

List of Figures

Figure 1	Countries with D4FL programme activities by partner	9
Figure 2	Survey respondents’ views on importance of potential FL data improvements.....	15
Figure 3	Distribution of D4FL resources and detail of evaluated D4FL resources	35
Figure 4	RAG-rated ToC.....	49
Figure 5	Original D4FL Programme Theory of Change	89
Figure 6	Updated D4FL Theory of Change	90
Figure 7	ACER UK ToC mapped onto the D4FL ToC.....	91
Figure 8	PAL Network ToC mapped onto the D4FL ToC	92
Figure 9	World Bank ALMA ToC Mapped onto the D4FL ToC	93
Figure 10	World Bank Accelerator ToC mapped onto the D4FL ToC.....	94
Figure 11	UNICEF’s ToC mapped onto D4FL ToC	95
Figure 12	NFER ToC mapped onto the D4FL ToC.....	96
Figure 13	Pyramid approach to engagement	133
Figure 14	Survey respondents per country	137
Figure 15	Biggest education priorities	138
Figure 16	Tools used to collect data on learning outcomes	139
Figure 17	Perceptions on the importance of reporting on SDG 4.1.1.....	142

List of Boxes

Box 1	Degrees of prioritisation of national needs.....	16
Box 2	D4FL contribution to improvements in learning data coverage and quality.....	27
Box 3	The case of Mexico: building bridges to use of citizen assessment data.....	29
Box 4	Support for the PAL Network and closing the gap on SDG 4.1.1a reporting	32
Box 5	In Ghana ownership, analysis, and use of data support likely impact	42
Box 6	In Somalia, a longer and more difficult road to impact is likely	42
Box 7	Country cases selection	111

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, Adele Rowe 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 middle-income countries (MICs) 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 Sustainable Development Goal (SDG) 4 numeracy and literacy reporting and the Group of Seven (G7) girls' education targets (*Track*), (ii) generate more and more quality data through learning assessment tools, support with technical assistance (TA), and promoting 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 Foreign, Commonwealth & Development Office (FCDO) 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 the Global Education Monitoring Report (GEM-R), the UNESCO Institute for Statistics (UIS), the World Bank, ACER UK, the People's Action for Learning (PAL) Network, the National Foundation for Educational Research (NFER), and the United Nations Children's Fund (UNICEF). The programme started operations in the United Kingdom (UK) Fiscal Year 2023/24 with a budget of GBP27 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) Global learning data gaps filled, including in emergency contexts; (2) More disaggregated learning data available that can be used to design and monitor foundational learning interventions; and (3) Increased capability in priority countries to use learning data to design and monitor foundational learning interventions.
- 4. The evaluation will cover all three strands of D4FL (*Track, Generate, Use*).** An external evaluation of the GEM-R was conducted in 2023; therefore, the terms of reference (TOR) excluded the *Track* component from the evaluation except for its coherence and interaction with the other D4FL components. In the inception phase, this exclusion was narrowed down to support for GEM-R and UIS, the major partners under the *Track* component. As a result, the mid-term evaluation (MTE) concentrates on five 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 the evaluation in October 2024, and UNICEF and NFER, who have been contracted since.
5. The evaluation team completed a formative evaluation in 2024/2025, and this MTE report is the first part of the summative evaluation, to be followed by a final summative evaluation in 2026/27.
- 6. The MTE aims to assess the extent to which – and how – the design and implementation of D4FL to date supports the achievement of its intended outcomes.** The overarching questions for the MTE are:
 - Is the D4FL programme producing the intended results?

- Is the D4FL programme representing good value for money?

7. **Data collection for the MTE was undertaken between 31 March and 05 November 2025 – midway through the D4FL implementation period.** The methodology 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, NFER), Mexico (PAL Network), Nepal (World Bank, PAL Network), Senegal (World Bank, PAL Network) and Somalia (World Bank, ACER UK).

8. **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 (six) that were possible given resources, a partial reliance on research with internal programme stakeholders and partner reports, limited response to some evaluation activities, and limited inclusion of countries who did not take up D4FL support. We countered these limitations by triangulating evidence and including participants not directly connected to D4FL. We note where there is less confidence about specific findings due to these limitations.

9. **This MTE report presents findings on implementation of D4FL to date** and makes recommendations to improve programme design and implementation for the remainder of the programme period.

Evaluation Findings

Relevance

10. **Global learning poverty persists, and the aim of the D4FL programme to contribute to progress against the SDG 4.1.1 foundational learning targets and learning poverty reduction remains as relevant for LICs and LMICs and the global education community as when the programme started.**

11. **Global learning data gaps also remain, and therefore the programme's *Track* objectives remain relevant.** The D4FL programme has responded well to the need of the global education community to demonstrate that countries can generate comparable global data by funding experimentation with different pathways to this goal. However, the programme has not yet funded much activity to support its goal of global learning data including coverage of emergency contexts, and its commitments to disaggregated data for children affected by other factors of exclusion like disability. Partners note they have very limited ability to shift existing country assessment practices in this regard, within the scope of their D4FL-funded work.

12. **D4FL's *Generate* and *Use* objectives are aligned with the needs of LICs and LMICs, which still often lack regular, representative and disaggregated data to monitor learning outcomes and the capacity to use this information to inform action on foundational learning.** Countries prioritise the generation of learning data for use in developing and tracking education policies and interventions over global reporting, but also see value in the ability to benchmark their education systems regionally and globally through SDG 4.1.1 reporting.

13. **With some exceptions, the D4FL partners have implemented the programmes in ways that respond to countries needs and the context of their education systems.**

Coherence

14. **Partners share a unified aim of improving learning outcomes and agree on the**

importance of quality foundational learning data to drive action towards this aim.

However, incoherencies and disagreements exist within the D4FL partnership, including around how much emphasis should be placed on global reporting vs. national use, and the value of non-government data on foundational learning. This lack of shared narrative risks confusing advocacy efforts and undermining the value of the data produced through D4FL activities in the future.

15. **Overall, D4FL demonstrates a high level of coherence with other global and national efforts to strengthen foundational learning data**, though this coherence is not yet fully consistent across all contexts. The programme is broadly seen as complementary rather than duplicative, aligning with existing national assessments and collaborating with key actors to support harmonised approaches.

16. **In several countries, D4FL has added value by filling gaps in technical capacity, promoting alignment with SDG 4.1.1 reporting requirements and contributing to increased coherency.** However, coherence is often dependent on strong local relationships and FCDO advisor facilitation, meaning coordination between partners and between D4FL and other FCDO-funded programmes varies across countries.

17. **There remains room for more systematic joining-up of D4FL activities and D4FL and other FCDO-funded programmes at country level.** Clearer pathways for how D4FL and related initiatives should complement one another, and more intentional action on this, could strengthen potential under the programme's *Use* objectives.

Effectiveness

18. **Available evidence indicates that the D4FL programme can deliver on its objective of filling global learning gaps**, through support for the generation of foundational learning data for 27 countries that can be submitted to UIS for vetting. In delivering this support, partners have used methodologies that can be replicated to track foundational learning elsewhere. There is, however, still uncertainty about whether the data will be accepted in all instances, and it has made little progress on filling gaps for children affected by emergencies and protracted crises.

19. **It will also deliver on its objective of more disaggregated data that can support policy and instruction, but not on its objective of building capabilities to use data.** The D4FL programme is likely to deliver on its objective of supporting more or better-quality disaggregated data that can be used to design and monitor foundational learning interventions, as D4FL support strengthens assessments and assessment capacities in countries. In a limited number of countries, D4FL-funded activities are likely to lead to data that are more easily accessible for use, or made useful for application and decision-making in country. The D4FL programme has not, however, strengthened capabilities to use data in any significant way. The vast majority of TA has focused on the capacity of technical individuals and/or departments to generate and analyse assessment. There are only two cases of D4FL support for building the capacity of broader education actors to use assessment data reported in D4FL partner reports (ALMA support in Guyana and Lao People's Democratic Republic (Lao PDR), and the evaluation has not yet had opportunity to explore the effectiveness of this support with country stakeholders. D4FL's contribution to filling country gaps on learning data for populations that are often excluded, is also likely to be limited, as this has not been made an intentional focus.

Efficiency

20. **Broadly, the D4FL programme's allocation of its resources, both by FCDO and by partners, is in line with its aim and objectives.** All the partner programmes funded contribute

to the D4FL aim, except for the Accelerator which while contributing to the programme's aim (reduced learning poverty) does not contribute substantially to its learning data tracking, generation and use objectives, as it is focused on interventions to improve foundational learning. The MTE identified three cases where Accelerator activities do contribute to generating data (Senegal) or to demand for the use of data (Cambodia and Rwanda).

21. The allocation of most resources to the *Track* strand could raise concern about its efficiency in delivering on addressing the ultimate aim of reducing learning poverty. This is because the causal pathway to in-country action to strengthen learning outcomes is longer and even more risky from global reporting than it is from data generated specifically for in-country use.

22. However, in most cases funded interventions that focus on global reporting are also contributing to strengthened national assessments. This strengthens the potential for impact and is an encouraging signal on cost-effectiveness.

23. The programme also continues to crowd in and leverage resources from other sources of funding, increasing efficiency towards its aims. The World Bank's ALMA and Accelerator programmes are funded from the Foundational Learning Compact (FLC) Umbrella Trust Fund which combines resources from multiple donors to support the Bank's foundational learning education programming. Meanwhile, accountable grant partners UNICEF and PAL Network have used co-funding arrangements to expand project scope.

24. Although programme implementation has faced delays, these are not a cause for concern as they are being managed and mitigated by partners and FCDO. The D4FL programme has been able to influence and monitor how partners are using resources, but this influence is weaker in the case of non-accountable grants.

Impact

25. There is little evaluation evidence on progress toward impact at this mid-term stage. However, there are some positive indications about likely progress toward impact:

- Capacity building of assessment authorities, which increases the potential for impact through strengthened capacity for future data generation and government ownership.
- In ten countries supported by ACER UK or the World Bank, activities that link data generation to data use are part of the scope of work, increasing the potential for impact through more and improved use of data to inform action.
- The PAL Network has developed detailed action plans to support the use of D4FL-generated data and to use D4FL-generated data to apply pressure for action on foundational learning
- The Accelerator Programme is directly working on addressing foundational learning poverty reduction through interventions in curriculum, materials and instruction.

Sustainability

26. There is evidence of promise regarding the sustainability of D4FL results at country- and global-level. Across many countries, foundations for sustainability are being laid through strong government ownership, the institutionalisation of national assessments, and capacity-building efforts that are strengthening local systems and organisations.

27. However, the evaluation has also identified notable risks to sustainability. This includes continued reliance on external funding and TA, particularly where item and test development, psychometric expertise, and costly software are required, turnover of key personnel,

particularly where TA focuses on individuals within government, and political economy dynamics – such as sensitivity to negative results or shifting priorities.

28. Overall, while D4FL has set important building blocks for sustained benefits, long-term sustainability will depend on institutionalised resourcing, distributed capacity (beyond individuals), and continued incentives for countries to value and use foundational learning data beyond the programme window.

Conclusions

Is the D4FL programme producing the intended results?

29. Outcome 1, “Global learning data gaps filled, including in emergency contexts”, will likely be only partially met¹. The programme will result in more internationally comparable data on foundational learning being generated. It is yet to be seen, however, whether data generated through or as a result of support from the D4FL programme will be accepted by UIS for SDG reporting. In addition, it is unlikely that the programme will make a notable difference to global learning data gaps for emergency contexts. Both limitations diminish the potential of the programme leading to impact on foundational learning outcomes for all children.

30. Outcome 2, “More disaggregated learning data available that can be used to design and monitor foundation learning interventions”, will likely be met. D4FL is expected to result in new, high quality foundational learning data being generated and in national assessments in a number of countries being strengthened. These data will be available for and have the potential to inform the design and monitoring of foundation learning interventions and policies, particularly where data use has directly informed the design of assessments and assessment systems.

31. Outcome 3, “Increased capability in priority countries to use learning data to design and monitor foundational learning interventions”, is at risk of not being met and, if it is met, will likely be only partially so. Only a minority of D4FL-funded support is directed at strengthening the capacity of data users within government, with TA activities largely focused on capacities for data generation. If capability to use learning data is not improved in priority countries, the intended pathways to impact of increased government action and improved quality of action on foundational learning in priority countries are at risk. As a result, the programme’s potential for impact is reduced.

Is the D4FL programme representing good value for money?

32. In many respects, the D4FL programme presents good value for money. The programme objectives are relevant to the global education sector and LICs and MICs. In most cases funded activities are likely to achieve their intended results efficiently and in a synergistic way: complementing and not duplicating other efforts. There are also positive signs of sustainability in a number of supported countries. In addition, the programme crowds in and leverages additional funding, furthering its aims.

33. Despite these strengths, several features of D4FL’s implementation reduce its value for money. Most notably, D4FL’s realisation of its *Use* objectives is lagging. Supporting capacity for data use is critical for D4FL-supported data generation and assessment-strengthening activities to realise their full potential, and therefore a lack of attention to use limits value for money. In some instances, global reporting has been prioritised over national use, further amplifying this risk. Other

¹ We RAG-rate the programme’s performance on the concluding questions, where red (R) equals not met, amber (A) equals partially met, and green (G) equals met.

risks to value for money include results not being sustained due to a lack of funding, a lack of political will, personnel turnover within supported governments, and gaps in what is left behind to replicate assessments; and risks of incoherent and competing activities and/or narratives across funded partners.

Recommendations

Recommendation 1. Deepen the focus on data for in-country use in all cases of country-level support.

34. **Rationale:** The D4FL programme has always contained potential trade-offs between the demands of generating globally comparable data for tracking purposes and of generating more, and better, data for use in countries to improve learning outcomes. While both can lead to action on learning outcomes at country level, the route to action from global reporting depends on the assumptions that countries will feel global pressure to improve learning outcomes, and that development partners will respond by funding interventions to improve outcomes. With changes in the global development context, where the sustainability of the global development and multilateral architecture is in doubt and where funding for interventions is scarce, D4FL should now be focused on ensuring that its support for data generation puts country use of the data at the centre.

35. **Action:** In the data-generation interventions that are still under way, D4FL partners should work to ensure that country assessment capacity and national use of data are at the centre of the support, even when the initial reason for the support is reporting against SDG 4.1.1. If any further country-level support is contracted, the scope of work should aim first to build country assessment capacity, with the ability to report globally as the by-product, where it is on the agenda.

Recommendation 2. The programme/FCDO should drive alternative options of supporting countries to build their assessment capacity.

36. **Rationale:** The D4FL-funded model of providing TA to build national governments' assessment capacity may not be fully replicable and sustainable in all LICs and LMICs requiring support, because of the cost and availability of TA. Conversely, other D4FL-funded models, such as the implementation of AMPLs and MICS, require less targeted TA but bring fewer benefits for country use of data. This requires new thinking about how support can be provided to help countries strengthen their data generation and data use capacities, while enabling global reporting, in a cost-effective and sustainable way.

37. **Action:** The D4FL programme/FCDO and its partners should consider what cost-effective options of support to strengthen foundational learning data generation (and use) could look like in future. This includes *what* and *how* support might be provided. Thinking about *what* means considering which capacities are most crucial, cost-effective and sustainable to strengthen within government, and which might be more efficient to outsource. Thinking about *how* means considering a smaller number of institutions to build capacity within. Options could include building regional hubs, as multi-government regional centres, private enterprises or non-government not-for-profit regional organisations. This would require in depth discussions on what capacity building would still be required to such centres, but might resolve cost and scaling issues regarding building capacity in all LMICs, capacity absorption and sustainability concerns.

Recommendation 3. Implementing partners should extend their efforts on ownership and use of data in their country-level support.

38. **Rationale:** The MTE has concluded that D4FL is still at risk of not meeting, or only partially meeting, its *Use* objectives. There is still time for D4FL, its partners and country team leads to build stronger links to data use, where such links are not present in its investments in data generation.

39. **Action:** Four related actions are proposed:

- Over the remainder of the programme, implementing partners should consider the extent to which they can work beyond technical assessment units and with broader education authorities to instil more ownership of data to support its use.
- In any new support for data generation, country needs for data to inform policies, programming and instruction should lead support for learning data generation.
- Any new support for data use should include more activities to support school managers and teachers to use assessment data.
- The D4FL programme team, other CMP teams and Education Advisors should work to maximise complementarity between the different strands of FCDO support, so that data generated with support from D4FL are used effectively.

Recommendation 4. The D4FL programme should seek ways to ensure that learning data are more inclusive, including more data for children affected by emergencies.

40. **Rationale:** The D4FL programme targets more global learning data for Education in Emergencies and Protracted Crises (EiEPC) settings as a primary programme objective. Data for other often excluded groups, such as children affected by disabilities, are also targeted. While acknowledging that D4FL partners have had limited ability to change whether often-excluded populations are included in the assessment they support, the MTE nonetheless concluded that D4FL will deliver on these objectives to a limited extent only.

41. **Action:** The formative evaluation recommendation stands that D4FL should increase its focus on the inclusion of children affected by emergencies. In addition, it should seek ways to help countries set up pathways to better availability of disaggregated data for children with disabilities and any other drivers of exclusion from assessment. This can be done by D4FL using its influence to advocate globally for urgent continued action on global standards and open-source tools to assess children affected by EiEPC and disability; the D4FL Programme Team arranging a workshop with the D4FL partners to discuss strategies for promoting more inclusive learning data on often-excluded children to countries; and the D4FL partners implementing the agreed strategies over the remaining programming period, and reporting to the D4FL Programme Team on how they have done so.

Recommendation 5. The D4FL programme and its implementing partners should focus on the sustainability of results in the remaining programming period.

42. **Rationale:** The evaluation shows that D4FL will achieve its objective of generating more disaggregated data that can be used to design and monitor foundational learning interventions and may also contribute significantly to filling global learning data gaps. However, whether these results will be sustained beyond the end of the programme is at risk, from several contextual factors that the D4FL programme can still help mitigate, as well as programmatic factors that are more under partners' control. There is still scope, in addition to what is already recommended in recommendations 1-4, to act to mitigate the risks and effects of some of these factors.

43. **Action:** In addition to the sustainability benefits from acting on recommendations 1 to 4, FCDO and its partners should aim to strengthen the following aspects in their support for assessments:

- budget institutionalisation of assessments
- building organisation-level capacity, rather than focusing support on motivated individuals
- ensuring that what is left behind to support replication of supported assessments is robust and comprehensive.

1. Introduction

1. This is the draft report of the mid-term evaluation (MTE) of the Data for Foundational Learning (D4FL) Centrally Managed Programme (CMP) of the Foreign, Commonwealth and Development Office (FCDO). The evaluation was commissioned by the FCDO. The terms of reference (TOR, see Annex 1) required two component evaluations to achieve its learning and accountability objectives, a formative and a summative evaluation. The summative evaluation was broken down further into a mid-term and final evaluation.

2. Mokoro Limited and Oxford MeasurEd are undertaking the evaluation between 2024 and 2027. The evaluation team completed a formative evaluation in 2025, and this report is the first part of the summative evaluation. It provides findings midway through D4FL implementation and makes recommendations to improve programme design and implementation. The MTE will be followed by the final summative evaluation in 2026/27.

1.1 Scope and purpose of the evaluation

3. The D4FL programme is centrally managed by the Girls' Education Department (GED) of the FCDO Education, Gender and Equality Directorate (EdGE). It succeeded the FCDO Better Education Statistics and Global Action to improve learning (BESTA) programme. D4FL has three strands of work: (i) *Track* global learning outcomes to enable numeracy and literacy reporting for Sustainable Development Goal (SDG) 4; (ii) *Generate* more data through learning assessment; and (iii) ensuring data *use* by policy makers and partners in the planning, design and implementation of foundational learning. It is implemented with: the UNESCO Institute for Statistics (UIS), the Global Education Monitoring Report (GEM-R), the World Bank, ACER UK, the People's Action for Learning (PAL) Network, the United Nations Children's Fund (UNICEF) and the National Foundation for Educational Research (NFER). The D4FL programme started in the United Kingdom (UK) Fiscal Year 2023/24 and will run to 2027. The indicative programme budget was GBP27 million, reduced to GBP25 million in 2025 after the UK Government announced a reduction to 0.3 percent in aid spending by 2027 and unallocated spending was cut from programme budgets.

4. The objective of our evaluation as set out in the TOR, is to assess the extent to which – and how – the design and implementation of D4FL supports the achievement of its intended outcomes: global learning data gaps filled, including in emergency contexts; more disaggregated learning data available that can be used to design and monitor foundational learning interventions; and increased capability in priority countries to use learning data to design and monitor foundational learning interventions.

5. The evaluation covers all three strands of D4FL (*Track*, *Generate* and *Use*), using all six OECD-DAC evaluation criteria (Relevance, Coherence, Effectiveness, Efficiency, Impact, and Sustainability). However, in terms of the TOR, the evaluation is not expected to evaluate the support to GEM-R and UIS in the *Track* strand in full, but to focus on its coherence and interaction with the other D4FL components. The MTE therefore concentrates on five D4FL partners: the World Bank, ACER UK, the PAL Network, UNICEF and the NFER. The latter two were contracted in 2024/25.

Mid-Term Evaluation

6. The MTE straddles learning and accountability objectives. Its overarching questions are:

- U

7. The evaluation questions (EQs, Table 2 below) are linked to the OECD-DAC evaluation criteria. We answer the EQs in the findings section of this report, and the overarching questions in

conclusion, setting out the reasons why the programme did or did not produce the intended results and represented good value for money. In the absence of sufficient, detailed financial data from all partners, we evaluate value for money qualitatively, by assessing whether the D4FL programme's objectives were achieved in mutually reinforcing ways that ensured optimal returns for D4FL, and whether the results will be sustained and lead to the intended impact.

1.2 Organisation of the evaluation report

8. The remainder of this evaluation report is organised as follows:

- **Chapter 2 – provides a summary of the context** for the D4FL programme and a **description of the D4FL programme**. This chapter is linked to Annex 2 (a timeline of key global events in learning assessment tracking), and Annex 3 (a detailed context for the D4FL programme). This chapter is also linked to Annex 4 which provides the original and updated D4FL Theory of Change (ToC) and explicated ToCs for each of the partners, and Annex 5 which is an overview of contracted work by implementation partner, and work delivered).
- **Chapter 3 – presents the methodology in summary**, with detailed EQs and approach to answering them. This chapter is linked to Annex 6 (which provides a detailed discussion of the methodology) and Annex 7 (which provides the detailed Evaluation Framework). Annex 8 presents the evaluation use and influence plan and progress against it.
- **Chapter 4 – presents the evaluation findings**. This chapter answers the EQs, by criterion. It is linked to Annex 9 (survey results) and Annex 10 (a table of the EQs, sub-questions and assessment criteria, and MTE judgements against these criteria). Annex 4 which presents the explicated ToC for the D4FL programme and its partners is also relevant to this chapter.
- **Chapter 5 – presents the Conclusions and Recommendations**. The explicated ToC for the D4FL programme in Annex 4 is relevant to the recommendations.

2. The Data for Foundational Learning Programme and its Context

2.1 Context

9. SDG 4 aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.² It marks the first truly global agreement in education on learning.³ One of the most notable changes in SDG 4 in comparison to previous education goals is its shift in focus from access to quality in education.⁴ SDG 4 emphasizes the importance of ensuring that students are learning once they are in the classroom. One of the major outcomes of this shift has been the heightened importance of robust and reliable assessment systems to evaluate learning.

10. 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 SDG 4, which can be seen as a compromise between economics-driven and human-rights-based perspectives (see paragraph 3-4 of Annex 3), emphasising quality and inclusivity as an overarching goal while including specific targets addressing each view.⁵

² UNESCO, 2015

³ UNESCO, 2017

⁴ Sayed and Moriarty, 2020

⁵ UNESCO, 2015

11. SDG 4 has been broken down into ten such targets, each with associated indicators, so that it can be monitored and assessed. 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. For SDG 4, the indicators chosen were influenced both by political aspects and by data availability. For example, target 4.1 represents a broad, 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”.⁶ However, the indicators for this target refer only to proficiency levels in math and reading and completion rates in formal schooling. 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.

12. Another challenge is defining proficiency levels (see paragraph 7 of Annex 3). 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.

13. Lastly, there are tensions around the choice of assessments to measure proficiency levels, even at the national level (see paragraphs 8-10 of Annex 3). Ideally, this decision should respond to a clearly defined purpose and use methods that are aligned with it and with resource availability.

14. 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 this view is increasingly questioned (see paragraphs 10-12 of Annex 3). ILSAs are expensive, time-consuming, complex and require significant technical capacity and expertise. This has led to calls for nimbler, cost-effective approaches aligned with the policy goals desired and with due consideration for timing and resource constraints.⁷ However, no one assessment can serve all purposes and misalignment of assessment approach and purpose generates stakeholder resistance.

15. Overall, we can see that while there is a formal consensus around SDG 4, 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 SDG 4 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.

Progress towards foundational learning according to current SDG 4 monitoring

16. Of the ten targets for SDG 4, 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

⁶ UNESCO, 2015

⁷ See for example Wagner et al., 2018

education.⁸ 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

17. SDG 4.1.1⁹ 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 minimum proficiency levels (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 2018.¹⁰ This effort resulted in the global descriptors for MPLs outlined in Table 1 below (see full table on MPLs mapped to different assessments in Annex 3, Table 7).

Table 1 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

18. 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.

19. The UN SDG Extended Report 2025 (as of 30 Apr 2025) reiterates that 4.1.1 reporting is from international/regional large-scale assessments (LSA) programmes (e.g., PISA, PIRLS, TIMSS, PASEC, LLECE, SACMEQ, SEA-PLM, PILNA). It notes end-primary and lower-secondary coverage remains stronger than in grades 2/3 and emphasizes UIS’s work on defining MPLs and linking.¹¹

20. According to the latest available data, approximately 70% of education systems have reported some data towards indicator 4.1.1, but only about 53% have enough data to be able to report trends. These percentages also vary widely across regions (see in Table 8 of Annex 3).¹²

⁸ UNESCO, 2015

⁹ 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.”

¹⁰ UNESCO, 2018

¹¹ UN, 2025

¹² Gustaffson, 2022

21. In September 2024 UIS highlighted a major infusion of 235,596 national data points (54% SDG indicators); which bolstered SDG 4 learning and related series that feed 4.1.1 reporting.¹³ In September 2025 UIS pushed a broad update to its education database (200+ countries) to support SDG 4 monitoring, with refreshed SDG 4.1 series accessible via the UIS Data Browser.¹⁴

22. At the same time the quality, reliability and MPL alignment 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 Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN (PASEC), and finally 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 considered unsuitable for reporting purposes for different reasons, including representativity, administration method and comparability.¹⁵ National assessment data suffers from similar issues.

23. In response to data quality, reliability and alignment with MPL, the UIS has specified a set of seven key reporting criteria, each containing a set of technical requirements that assessments must meet to be eligible to report against SDG 4.1.1. The document has been created to support the application of reporting criterion 1 – Alignment to the MPL and construct validity, by providing a clear definition of the MPLs and a blueprint for each educational stage – a) at the end of lower primary (Grades 2/3), b) at the end of primary, and c) at the end of lower secondary.¹⁶

Progress towards SDG 4.1.1 according to available data

24. Regarding progress, most countries are not on track to achieve SDG 4 by the target year of 2030. According to estimates of the Inter-agency and Expert Group on SDG Indicators (IAEG-SDG) in 2023, even if trends reported before the pandemic remained unchanged, only one in six 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.¹⁷ 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.¹⁸ However, these advancements have been slowing down consistently. In addition to this, COVID-19 has had a profound impact (see Annex 3 paragraph 24).

25. The 2025 Sustainable Development Goals report showed most countries are off track to meet targets for access, completion and learning outcomes. UIS emphasizes that **many countries are now reporting**, but **trend lines mostly show flat or declining proficiency** once harmonized to the Global Proficiency Framework (GPF). Fewer than **half of students worldwide** reach the **minimum proficiency level** in end-of-primary reading or math (Indicator 4.1.1 b). Early-grade learning (4.1.1a) remains weakest — foundational literacy and numeracy gains have stalled.¹⁹

¹³ UNESCO, 2024

¹⁴ UNESCO, 2025

¹⁵ Montoya, 2024

¹⁶ UIS and ACER UK, 2025

¹⁷ United Nations Economic and Social Council, 2023

¹⁸ Gustaffson, 2022

¹⁹ UIS and ACER UK, 2025

Pending challenges

26. Annex 3 provides more detail on several continuing challenges.
27. **Exclusion:** The phrasing of indicator 4.1.1, which measures the share of children in school achieving MPLs, does not account for those out of school. Consequently, the indicator fails to capture the full picture of educational disparities and challenges (see Annex 3 from paragraph 26).
28. **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 3 from paragraph 30). Despite these challenges, there is evidence of assessments influencing policy, albeit mostly through top-down changes. Areas most influenced tend to be curricula and monitoring, evaluation and learning systems. Changes to financing, pedagogy, or teacher training are much rarer.
29. **Misinterpretation of data:** Media coverage of learning outcomes can be hugely influential, but difficulty in interpreting results often leads to media outlets misrepresenting reports, leading to reactions that do not align with the actual findings (Annex 3 from paragraph 33). This misalignment between the intended assessment messages and public perception can hinder effective use of data.
30. **Sustainability and affordability:** Many of the assessments deemed valid for reporting are both expensive and complex to implement, an issue that affect fragile states and regions experiencing educational emergencies disproportionately. Recent and ongoing funding cuts in the sector will leave gaps to sustain and improve. Efforts to build national and regional measurement capacity need to ensure affordable and practical assessment, and strengthen system capacity, for more reliable and actionable learning data (Annex 3 from paragraph 36).

2.2 The Data for Foundational Learning Programme

Programme design and theory

31. The D4FL programme ToC, as adjusted after the formative evaluation, sets out the programme theory to deliver the following high-level objectives.
- **Impact:** Contribute to progress against SDG 4.1.1 and learning poverty reduction in LICs/MICs.²⁰
 - **Outcome 1:** Global learning data gaps filled, including in emergency contexts.
 - **Outcome 2:** More disaggregated learning data available that can be used to design and monitor foundational learning interventions.
 - **Outcome 3:** Increased capability in priority countries to use learning data to design and monitor foundational learning interventions.
32. The achievement of D4FL's outcome objectives is expected to contribute to its impact, through increased pressure on action for learning, increased action and improved quality of action on foundational learning in priority countries. The original programme ToC is shown in Annex 4 Figure 5 and the revised ToC in Figure 6. It shows how the D4FL programme is intended to achieve these high-level objectives through three strands of work: *Track*, *Generate* and *Use*.

²⁰ SDG 4.1.1 refers to the 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. Learning poverty measures the share of children who have not achieved minimum reading proficiency by age 10, adjusted by the proportion of children who are out of school (and are assumed not able to read proficiently).

33. The *Track* strand focuses on global tracking of learning outcomes and specifically enabling LICs and MICs to report against SDG 4 foundational learning targets and the Group of Seven (G7) girls' education targets. In the revised ToC the *Track* strand comprises three intended outputs and associated activities:

- **Output 1:** Coordination and publication of comparable learning data using quality assessments
- **Output 2:** Development and implementation of global methodologies to track foundational learning for all internationally
- **Output 3:** Generation of comparable learning data using quality assessments.

34. Planned activities to achieve Output 1 are funding the GEM-R to publish global reports and funding the UIS and UNICEF to track learning data. Output 2 is to be achieved through funding UIS and UNICEF to develop methodologies and implement global public goods to close learning gaps, and the PAL Network to develop and implement internationally comparable citizen-led assessment; Output 3 through generating data using this citizen-led assessment, as well as from the results of Output 4 (adaptation or implementation of rigorous regular primary learning assessments). The achievement of Outputs 1 to 3 is expected to lead to Outcome 1 through more LICs and MICs reporting using comparable and disaggregated learning data, including in emergency contexts.

35. The *Generate* strand is focused on supporting LIC/LMIC governments, with one output:

- **Output 4:** Adaption and/or implementation of rigorous regular primary learning assessments.

36. Activities to achieve Output 4, are funding TA to add/strengthen quality learning components to existing national data systems. Two implementing partners contribute to this strand: ACER UK and the World Bank (through the Accelerating Learning Measurement for Action (ALMA) programme).

37. Output 4 is expected to lead to Outcome 2 through strengthened national assessment systems / data in participating countries. The achievement of Output 4 is also expected to contribute to Output 3 and more LICs/LMICs reporting internationally, as it is expected that some countries would be supported to strengthen and/or link existing learning data so it could be used to report.

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

- **Output 5:** Strengthened mechanisms for use of learning data in national decision-making, policies and reforms.
- **Output 6:** Strengthened use of simple foundational learning assessment in classroom and community interventions.
- **Output 7:** Strengthened curricula, teacher development, pedagogy and teaching materials to support foundational learning.

39. Output 5 is also expected to flow from funding TA to add/strengthen a quality learning component in national data systems, and funding target setting and planning to improve learning for all, informed by data and evidence. This input is also expected to lead to Output 6. Support to FCDO bilateral programmes and CMPs to ensure quality learning data is applied in interventions to support instruction, was planned to support Outputs 6 and 7. Output 7 was also expected to be achieved through funding the World Bank Accelerator Programme to strengthen World Bank-funded interventions in countries' instructional core. In the absence of partners and activities funded to deliver on Output 6 the formative evaluation recommended a focus on teachers as a key stakeholder group for D4FL's *Generate* and *Use* activities.

40. The achievement of Outputs 5 and 6 was expected to lead to Outcome 3 through policies, targets and interventions in participating countries using learning assessment and data to strengthen curriculum, assessment, and instruction. The achievement of Output 5 was also expected to contribute to Outcome 2. Output 7 was expected to contribute directly to D4FL's impact goal.

Programme implementation

41. Programme implementation commenced in 2023/24, comprising *Track* activities with UIS, GEM-R and UNGEI/Cambridge Education. In 2023/24 further implementing partners were mobilised in the *Track*, *Generate* and *Use* strands, ACER UK, the PAL Network and the ALMA and Accelerator 2 Programmes of the World Bank Foundational Learning Compact (FLC) Umbrella Trust Fund. In 2024/25 UNICEF and the NFER were added as implementing partners.

42. By October 2025, the D4FL programme included activities (planned or already under way) at the global level, at country level in 30 LICs and LMICs, in two regional interventions (in the Eastern Caribbean and East Africa) and support for the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA). Annex 5 Table 9 provides an overview of all the D4FL partners, their contracted and agreed programme activities, and implementation progress. Figure 1 below provides an overview of all countries in which the D4FL programme is active, by partner.

3. Evaluation Methodology and Limitations

43. This MTE began on 3 March 2025 with data collection tool development, after agreeing the MTE plan in February with the FCDO D4FL Steering Committee managing the evaluation. Data collection and analysis occurred between 31 March and 5 November. The evaluation team worked independently and without interference to produce the report’s findings and conclusions. Draft recommendations were discussed with FCDO and partners, and we adjusted their focus and scope to ensure relevance and feasibility. Annex 6 describes the methodology, summarised below, in detail including ethical considerations. There is no known conflict of interest concerning any team members.

3.1 Overview of the approach

44. This evaluation is a mixed-method, realist, ToC-based evaluation. Following consultations with FCDO and implementing partners in the inception period, the team concluded that the existing ToC stood in principle, but worked with FCDO and partners in the formative evaluation to expand the ToC, so that the summative evaluation could test a more detailed version (see Annex 4).

45. The MTE included desk-based research, primary data collection and analysis at portfolio and country levels. The country-level work focused on the *Generate* and *Use* strands of the D4FL programme, but considered coherence and interaction between these and the *Track* strand as relevant.

3.2 Methodology

Evaluation Questions

46. In Table 2 below we outline the 12 high-level evaluation questions for the MTE, mapped onto the evaluation criteria and aligned to the explicated programme ToC. The MTE questions and evaluation framework were agreed with the D4FL Steering Committee, through the MTE evaluation plan. The questions are as in the inception report (IR), except (i) for language changes to reflect the ongoing nature of the D4FL programme in the MTE or to make the questions clearer; (ii) rearranging the coherence questions; and (iii) restructuring the effectiveness questions to reflect ToC changes.

Table 2 Summative evaluation criteria and questions (MTE)

Criteria	Questions
Relevance	1. To what extent does D4FL address the needs of LICs/LMICs?
	2. To what extent do D4FL’s objectives and activities address the needs of the global community?
Coherence	3. To what extent is D4FL a coherent programme across strands and partners?
	4. To what extent are D4FL activities coherent and complementary with other global learning data efforts?
	5. To what extent is D4FL coherent and complementary with other FCDO programmes?
Effectiveness	6. To what extent is D4FL leading to the availability of more disaggregated learning data in LICs and MICs that can be used to design and monitor foundational learning interventions?
	7. To what extent is D4FL leading to better capability in priority countries to use learning data to design and monitor foundational learning interventions?
	8. To what extent and how is D4FL filling global learning data gaps?

Criteria	Questions
Efficiency	9. How well is D4FL allocating resources to maximise the chance that outcomes are achieved?
	10. To what extent are partners delivering their activities in an efficient way?
Impact	11. What evidence is there that D4FL's results will lead to improved quality of education and reduced learning poverty?
Sustainability	12. Are in-country benefits likely to be sustained?
	13. What long-term changes in the global assessment landscape are likely to result from D4FL?

Data collection

47. In Annex 7 we provide the agreed evaluation framework for the MTE, outlining sub-questions, assessment criteria and data sources for each EQ. In Annex 6 we provide further detail on the nature and purpose of each data collection activity. Table 11 in Annex 6 provides an overview of the data collection design for the MTE and indicates how we used qualitative and quantitative data sources at country and global levels to address the evaluation criteria and questions included in Table 2 above.

48. At global level, we undertook 11 key informant interviews (KIIs) with FCDO (1), implementing partners (7) and independent stakeholders (3), as well as ten workshops with the FCDO and implementing partners (see Table 13 in Annex 6 for more detail).

49. At country level we selected six countries to study the implementation of the programme partners. The country cases were aimed at collecting evidence across countries on the implementation of the D4FL programme against the evaluation criteria in different contexts, rather than individual case studies in their own right. This design of the country case component of the methodology, aimed to cover more cases of D4FL-funded interventions across countries, rather than go more in depth in any one country.

50. The country cases covered four implementing partners, different regions, countries where there is/is not an FCDO education advisor, and countries in different stages of implementation (see Table 12 in Annex 6). We took the five cases studied in the formative evaluation forward to the MTE and added a country, allowing us to research the partners' work in the same country over time. The countries agreed with the FCDO were: Ghana (World Bank, ACER UK), Kenya (PAL Network, NFER), Mexico (PAL Network), Nepal (World Bank, PAL Network), Senegal (World Bank, PAL Network) and Somalia (World Bank, ACER UK).

51. We visited two countries in person (Ghana and Kenya), while data collection for the other four countries was done remotely. Across the six countries we undertook 3 KIIs with implementing partner representatives (14 KIIs), government (16) and other stakeholders (6), and FCDO advisors (3). Table 13 in Annex 6 provides more detail. For non-case countries we collected evidence through interviews with the FCDO and partners, partners' D4FL reports and the survey.

52. The team surveyed country education authorities and other education stakeholders in D4FL-supported and non-D4FL supported countries. We distributed the survey through the D4FL partners and multiple other channels.²¹ We had contacted several other potential distributors but without

²¹ Including FCDO Education Advisors, the FCDO What Works Hub, the PAL Network, ACER UK, the World Bank Task Team Leaders receiving D4FL support, GEM-R, and the OECD Education and Skills Directorate. The survey was also circulated during the Association for the Development of Education in Africa Triennale Conference in Accra, Ghana, by the African Development Bank.

success. Despite these efforts, the overall number of responses was limited,²² with 63 responses in total, and uneven coverage across countries, i.e. responses were received for 15 countries out of 30 where D4FL has been active through partners, and for an additional 12 countries. We used survey data only where supported by other evidence. A summary of the survey and its results is provided in Annex 9. Due to the distribution methods, a response rate for the survey cannot be estimated.

3.3 MTE Methodological Limitations

53. **Confidence about the contribution of D4FL to learning assessment outcomes:** The evaluation has distinguished the contributions made by D4FL and D4FL-financed activities within partners' programmes, from contributions by other stakeholders and other partner activities to the extent possible. The evaluation incorporated questions about other contributions to results in data collection. Because programme results were in almost all cases still at the output level, the team did not undertake full contribution analysis in the MTE. Even so, in developing the findings, the team considered the contributions, positive or negative, of other actors. Overall, its analysis of the D4FL programme's likely contribution to the programme's outcomes, is qualitative and about contribution rather than attribution, as it is not based on counterfactuals or statistical analysis.

54. **Potential bias in D4FL and partner documentation, and respondent bias in interviews and surveys:** 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 challenges experienced. The evaluation team mitigated 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 benefitting from programme resources, or directly benefitting from programme inputs, such as members of education groups at country level.

55. **Limitations in drawing evidence from the country cases** included the following:

- Only limited resources were available for each case, limiting the number of KIIs and options for collecting data directly with marginalised beneficiary groups.
- Country cases may not be representative of D4FL implementation. To mitigate, we also considered non-case country data from D4FL programmatic reports, global KIIs and the survey.
- In almost all countries D4FL activities were mid-way or nascent, limiting evidence on achieved effectiveness, impact and sustainability. We mitigated this by evaluating likely effectiveness, impact, and sustainability utilising early signals of progress (or lack thereof). At this stage, we are therefore also limited in explaining how activities contributed to objectives, and in undertaking contribution analysis.
- The D4FL programme itself is not visible as the funder of the implementation activities of implementing partners, which hampered access to respondents in cases such as Kenya, as they did not automatically see the evaluation as relevant to them. This lack of visibility also made it difficult for country stakeholders to distinguish D4FL activities from other activities. In mitigation, we carefully distinguished D4FL-supported activities from partners' other activities during questioning, excluded unrelated information unless it provided information on D4FL-supported activities (e.g. clarified D4FL's coherence or effectiveness), and cross-checked data between sources.

²² The team are unable to say what the response rate was, as it is unclear how many people were sent the survey.

56. Case study limitations and evidence gaps are acknowledged in reporting. The team will consider these lessons when developing the 2026/27 Final Summative Evaluation Plan in February 2026.
57. **Limited inclusion of countries who do not take up D4FL support:** The MTE survey included countries not targeted by D4FL, but respondents from only 12 countries out of 99 LICs and MICs not benefitting from D4FL-supported activities responded. We will consider further how to collect non-D4FL country perspectives in the final summative round, within the resources available.
58. **Low response rate to evaluation survey.** However, only 63 responses from 29 countries were received, including responses from 15 of the 30 countries in which D4FL-supported activities are implement. Survey results in the MTE may therefore not be fully representative of stakeholders beyond some of the D4FL-supported countries. To mitigate this, the team has triangulated survey findings with data from key informant interviews and desk-based research.
59. **Sources of data:** Some reporting of activities, outputs and outcomes relied on partners' reports. The evaluation team has reported transparently on information sources and any limitations.
60. **Costing of this multi-year evaluation.** The risk of costings increasing over the coming years to accommodate the scope of work required for the evaluation coverage is likely, but efficiencies achieved during the formative evaluation and data collection of the MTE have enabled the reallocation of resources which should help act as contingency and limit a significant change in the total value of the contract required over the remaining evaluation period.

4. Evaluation findings

4.1 Relevance

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

Summary of findings on relevance: Global learning poverty remains high and D4FL's focus on progress toward SDG 4.1.1 foundational learning targets remains relevant for LICs/LMICs, and the global education community. Persistent global learning data gaps also justify the programme's *Track* objectives. D4FL has responded to the sector's need to show that countries can produce comparable global data by funding experimentation with multiple pathways. However, it has funded limited work on generating learning data in emergency contexts, reflecting limitations in partners' ability to do so within D4FL's scope of work. The *Generate* and *Use* objectives align with LIC and LMIC needs, as many still lack regular, representative, and disaggregated data – and the capacity to use it – for monitoring and action. While countries prioritise learning data for policy design and tracking, they also value SDG 4.1.1 benchmarking. Going forward, balancing investment in global reporting with investment in data systems that serve countries' own evidence needs will be essential. Implementing partners have been largely responsive to country contexts.

EQ1 To what extent does D4FL address the needs of LICs/LMICs?

Finding 1. D4FL's goal of reducing learning poverty in low- and lower-middle income countries remains highly relevant. Its objectives of more disaggregated learning data, and more capability to use the data are also relevant to these countries.

Relevance of the D4FL goal of reducing learning poverty

62. Global learning poverty persists, and the aim of the D4FL programme to contribute to progress on the SDG 4.1.1 foundational learning targets remains as relevant as when the programme started. The GEM-R team estimates that 55 percent of children in low-income countries reach the end of primary school, but only 9 percent will have basic reading skills. In lower-middle income countries 86 percent reach the last grade, but 40 percent will not have learned to read.²³

Relevance of the D4FL objective of generating more disaggregated learning data

63. Many countries do not have regular learning data. Only 57 percent of the 124 countries in UNICEF's 2025 Foundational Learning Action Tracker (FLAT) survey undertook representative large-scale assessment of early grade reading and/or mathematics every two years or more frequently.²⁴ Moreover, only 44 percent used assessment data in classrooms and 49 percent provided support to teachers on foundational assessment. As said in one KII: "Many countries do not report on student achievement. There is a lack of awareness of the dimensions of the problem. Policymakers across the world...are generally not aware of how bad achievement is in their countries."

64. Generation of data was a priority in five out of six of the case study countries. Government KIIs in Ghana, Somalia, Nepal, and Senegal thought strengthened assessment important, because of gaps in data and/or capacities. In Ghana and Somalia assessments are an education strategic plan priority. In Mexico, where D4FL supports the PAL Network, with government learning assessments on hold alternative learning data collection was important, according to country KIIs.

65. The evaluation survey confirms the relevance of D4FL investing in data generation. Respondents were asked to rate the importance of improving learning data for government in their country zero to five. While only 64 percent of respondents selected four or five, only one of those selecting three or lower thought it was because the country already had enough data or it was not a priority. The remainder of those selecting three or lower, cited reasons related to lack of political will and incentives most often (90 percent of these respondents), followed by weak systems and poor assessment governance (64 percent) and a poor culture of and capacities for data use, and lack of resources (46 percent each). Arguably, these reasons confirm the relevance of the D4FL programme.

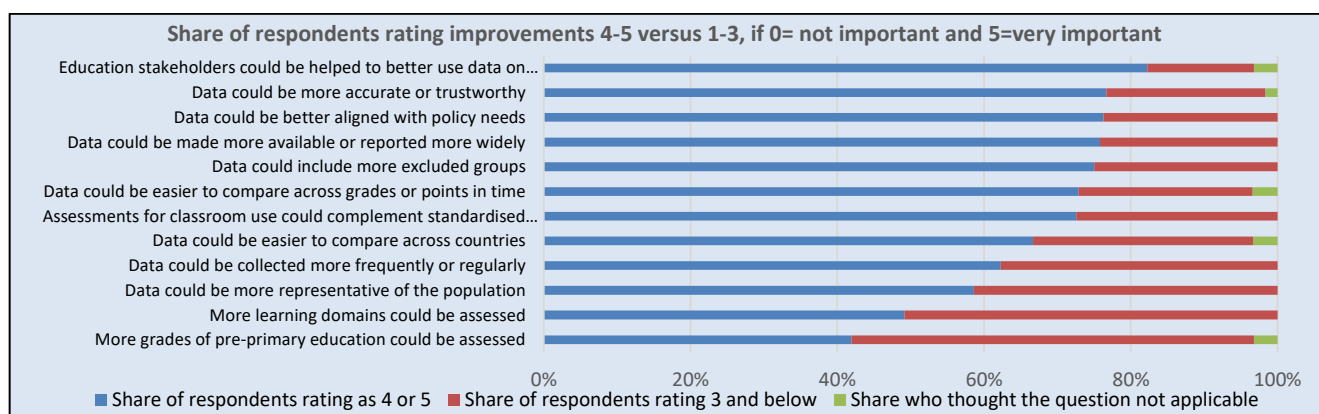
Relevance of the D4FL objective of more capability to use data

66. Government respondents in Ghana and Kenya said improved use of learning data was important. 84 percent of survey respondents thought that helping education stakeholders use foundational learning data was an important improvement for foundational learning data in their country. Other improvements seen as important by two thirds or more of respondents, were also linked to data use (e.g. data can be better aligned to policy needs and data could be made more available or reported more widely).

²³ UNESCO and GEMR, 2024

²⁴ UNICEF, 2025, 2025 FLAT Survey report, [link](#).

Figure 2 Survey respondents’ views on importance of potential FL data improvements



Source: D4FL programme evaluation survey

67. Classroom use of assessments emerged as a priority need in Somalia, Ghana and Senegal. In Somalia, teachers’ use of assessments was the top priority for the nascent assessment system in the national assessment framework. In Ghana, Government and donor representatives emphasised the need for more school-level stakeholder capacity to use data from Ghana’s census-based national assessments. In Senegal, the 2024 UNESCO Spotlight Report noted the need for classroom assessment.²⁵ Over 70 percent of survey respondents rated improving classroom data important.

Finding 2. Reporting against SDG 4.1.1 is seen as important for LICs/LMICs, but not as important as generating data that are aligned to policy needs and can be used to improve learning outcomes.

68. The take-up of D4FL’s support to help countries align with the UIS criteria for reporting against SDG 4.1.1, reflects countries’ interest in reporting foundational learning outcomes. Since the start of the programme, five country authorities have accessed ACER UK support for this purpose. A further eight countries, and one regional assessment authority, have accessed World Bank ALMA support. The evaluation survey confirms this interest, with 94 percent of government respondents rating reporting on SDG 4.1.1a (foundational learning) as important for their country. Country authorities in the country cases saw global reporting as beneficial because it signalled participation in the global order and benchmarked their education system against global learning standards.

69. Countries nonetheless give primacy to ensuring assessments fulfil national needs. In Ghana and Kenya government respondents considered data for national use the primary purpose of their assessment systems. ACER UK reported that in Malawi, country authorities were foregrounding national utility over global reporting, even though the initial support request was tied to global reporting and the Government still intend to use their national assessment for global reporting. The evaluation survey also signals that easier comparison of learning data with other countries is not a top priority for countries (see Figure 2 above) relative to other potential improvements to assessment and learning data.

70. Almost all global foundational learning stakeholders interviewed this year, including those funded by the D4FL programme to support global reporting, said the generation of data for use and the development of capacity to use data should be the focus in countries. As said by one: “Global reporting is acknowledged as a mechanism to ensure data quality, but it is not an end in itself.

²⁵ UNESCO, 2024

Countries should focus on building capacity to use assessments meaningfully.” Comparable global data is seen as a by-product of strong national systems, not the primary goal.

Finding 3. Getting the balance right between generating data for global reporting and strengthening national assessments and systems, is critical for D4FL’s ability to address global learning poverty, especially in the changing development context.

71. Global learning stakeholders interviewed were concerned that focusing heavily on global requirements risks diverting time, resources, and political capital away from national improvement agendas, particularly in lower-capacity systems and given shrinking development resources. Global reporting of learning data could prompt countries to improve foundational learning, but this would require global reporting leading to pressure and funding for action. Even then assessments done solely with the intent to report will be of limited use to guide such action.

72. At country level, partners recognise and discuss different assessment design decisions depending on whether the purpose is national decision-making or SDG 4.1.1 reporting, particularly relating to sampling, disaggregation, comparability, and which learners are included. The evidence shows that choices about assessment design shift depending on whether the priority is national use, pedagogical utility, or global reporting. Examples raised by or discussed with partners include:

- The design of samples (who is assessed and why) varies depending on whether the aim is international comparability or local decision support.
- Disaggregation for SDG reporting is structured by agreed indicators, whereas national systems may seek different breakdowns (e.g. by region, school type, marginalised learners).
- The design of what is assessed also varies. For example, countries differ (in both their written and enacted curriculum) in their focus on certain pedagogical approaches (like phonics) and some country approaches may not align with global frameworks.

73. Multiple evaluation participants acknowledged that design decisions for SDG reporting can reduce usefulness for instruction or policy without adaptation. Participants noted – and we observed – that data generation efforts can then become decoupled from data use, with increased risk that the data generated will not meet evidence users’ needs. We observed a spectrum of interventions ranging from TA building national assessment systems that inform policy-making, teaching and learning and *include* meeting the SDG 4.1.1 reporting criteria, through to data generation initiatives focusing *primarily* on SDG 4.1.1 compliance and global comparability. This is illustrated in Box 1.

Box 1 Degrees of prioritisation of national needs

Country case A: Ghana, prioritising national needs

In Ghana, D4FL supported the adaptation of the National Standardised Tests (NSTs) Primary 4 and 6 to be used to report against SDG 4.1.1a and b. ACER UK provided this support to the national assessment authority of Ghana in a hands-on manner responding to country needs, including improved comparability between the previous round of Primary 4 assessment and the round supported under D4FL. The support also comprised capacity building of the assessment authority, e.g. on analysis of assessment data for use.

Country case B: Malawi, prioritising SDG 4.1.1 reporting and national needs

In Malawi, the Government requested ACER UK to define a common vision across the Ministry on a system of assessments to support improving foundational learning for all. Consultation identified three priority purposes for assessment: (i) at least one assessment should aim at SDG reporting; (ii) a harmonised set of national and district assessments should aim at driving community, district and national resource allocation decisions for improved learning results; and (iii) a set of formative assessments should inform remediation for struggling early-grade students. ACER UK will be supporting (i) and (ii) under the D4FL grant. According

to ACER UK respondents, after further consultation Malawi is amending (i) to foreground national utility over global reporting.

Country case C: Nepal: prioritising global reporting needs

In Nepal, the World Bank ALMA work had aimed to embed AMPL into the national assessment process. However, delays – in part attributed to the Government prioritising national assessments – mean that this embedding may not be possible. AMPL was still seen as important as it would enable UIS reporting, and at the time of the case study the indication was that AMPL would be rolled out as a separate initiative if embedding/aligning with national assessments were not possible.

Country case D: Somalia: prioritising global reporting needs

In Somalia, the intent was initially to implement an AMPL in parallel with a national assessment, but after government concerns about the complexity of the parallel yet integrated exercise, AMPL implementation was separated from carrying out a national assessment. Government respondents indicated to the evaluation a desire to focus on developing national assessment capacity, even if interested in benchmarking learning against global standards and using the AMPL results to report.

Source: ACER UK Quarterly Reports and interviews, World Bank ALMA reports, case study interviews Nepal and Somalia.

74. Assessments designed primarily for global reporting have limited potential to build sustained demand for data use, as they often do not provide the granularity, timeliness, or actionable insights needed for local decision-making.

75. The D4FL programme includes promising models that balance strengthening national assessments for country use with readying assessments and data for global reporting. By count (ten countries and one regional assessment authority) support offered to countries under D4FL has more often aimed to integrate global reporting requirements into existing national assessment systems than to fund parallel assessments (four countries). This indicates that D4FL is managing this tension. However, given that trade-offs are present, the risk that the push for global reporting results in assessment designs that prioritise comparability over contextual relevance or utility for teaching and policy, must continue to be managed.

Finding 4. D4FL’s activities through the World Bank and ACER UK respond for the most part to the needs of supported country authorities. Support through NFER has the potential to do so.

76. ACER UK and the World Bank ALMA and Accelerator support respond to articulated partner country needs and work closely with country partners, for the most part.

- ACER UK’s demand-led D4FL operating model, its expertise and the way in which it functions in country, make it a relevant partner responding to country needs. It has adjusted its offering following a desk review of the country context, and tailored the timing of interventions to fit country assessment cycles. It has been responsive to specific requests of partner countries.
- The World Bank’s ALMA and Accelerator programmes respond to needs in the implementation of World Bank foundational learning programmes as identified by programme task team leaders (TTLs). In-country stakeholders see the World Bank as a trusted long-term partner. There is evidence that the World Bank adjusts its D4FL-supported work programmes to respond to in-country context, e.g. in Ghana and Senegal. However, we have also observed that support for parallel assessments proceed without full government buy-in (e.g. Nepal).

77. The NFER’s support to the Kenya National Educational Assessments Council (KNEC) for establishing the Education Assessment Resource Centre (EARC) is a highly relevant service pathway

to meet the needs of LICs/LMICs, establishing assessment capacity at the regional level. The Kenya case study confirmed that the NFER support was responsive to KNEC's needs and constraints.

78. UNICEF MICS is a useful complement where countries do not have quality data, but is of limited use for in-country policy and programming needs. UNICEF itself noted that the inclusion of the learning module in MICS was primarily to provide cross-country comparable data for reporting on the SDGs. In reality, MICS could not feasibly expand its learning data module to be of higher utility for in-country use, given its remit of collecting data on multiple household indicators.

Finding 5. The PAL Network may not respond directly to the needs of national education authorities, but its activities are nonetheless relevant to LIC/LMIC needs.

79. The relevance of the PAL Network for D4FL's objectives firstly stems from its ability to undertake evidence-based advocacy for better learning outcomes for all. In the past, the PAL Network has used its data to shift the debate to learning at country and global level, contributing to increased prominence of learning in country plans and in the international agenda.²⁶ For its D4FL-supported data round of ICAN/ICAR assessments, it has a strategy in place to advocate for learning interventions in countries and at regional level. The Mexico country case study for this evaluation presented evidence that country-level partners of the network use the data to design their own education interventions. The Senegal country case, however, offers evidence that a result chain from citizen-based data to country action can be difficult to build. In this case, the government stakeholders consulted showed little knowledge of or interest in the PAL Network data.

80. Secondly, the PAL Network data can complement official data, and can be the only nationally representative data. Global stakeholders thought D4FL's support for the PAL Network was relevant because "tools like MICS and PAL have value particularly where national systems are weak". These views are reflected in an earlier evaluation of the PAL Network, which found the provision of data inclusive of out-of-school children (OOSC) an important benefit of the Network's household-based surveys.²⁷ The Mexico country case, where the dissolution of standardised national assessments left a gap that the PAL Network fills for teachers, schools and regional governments, demonstrates this.

81. Thirdly, D4FL support has enabled the PAL Network to integrate the UIS criteria into its ICAN/ICAR assessments. This according to global stakeholders demonstrated "that reportable data can be generated through multiple service pathways". However, UIS has clarified that the choice of data to use for global SDG4 reporting, is governments' prerogative, even if the PAL Network's ICAN/ICAR assessments were well advanced on technical compliance with the UIS criteria.²⁸ This means that ICAN/ICAR data will only be used for global reporting when governments decide to do so. National officials from seven countries attended the ICAN/ICAR launch in Nairobi in November 2025, signalling that some governments may do so.

²⁶ R4D Institute, 2015

²⁷ Ibid

²⁸ UIS, 2025c

Finding 6. Across partners, D4FL still has only very limited activities to make progress on more inclusive data, including data on children in emergencies, despite this being a programme objective.

82. The D4FL programme explicitly targets more foundational learning data on children affected by emergencies, and data that can be disaggregated by other groups of children often excluded from assessment, such as girls, OOSC, children with disabilities, and children in rural areas.²⁹

83. Programme stakeholders agree this is an important objective. Most respondents to the evaluation survey saw more inclusive data as an important foundational learning data improvement (see Figure 2). Government respondents in Ghana, Nepal and Somalia thought data should be more inclusive respectively of children with disabilities, children in remote rural areas, and children in emergency schooling. D4FL partners acknowledge that there are gaps in the inclusivity of global assessment tools, especially on disability and learning in EiEPC settings. Annex 3 from paragraph 26 provides a brief overview of the challenges associated with exclusion in learning assessments.

84. In implementation, however, D4FL-supported interventions have targeted more inclusive assessment data that can be disaggregated only to a limited extent. In D4FL's support to national assessments, the extent to which the data were inclusive and could be disaggregated depended on how often excluded populations were already included in assessments supported by D4FL, as not one country prioritised more inclusive assessments for this support, despite it — at least in the country cases — being seen as important over the longer term.

85. D4FL support for the PAL Network and UNICEF MICS to integrate the UIS criteria is a positive contribution to the availability of more inclusive globally comparable learning data for all. Both are household surveys catching populations excluded from school-based surveys, such as OOSC and rural children, to provide a more accurate picture of learning. However, they still may not fully include populations affected by other drivers of exclusion, such as disability.³⁰

86. ACER UK's work with UNWRA is the first example of D4FL support for data generation on displaced children in emergency education. While there are other emergency contexts where the D4FL programme provides support – e.g. Somalia and Mali – in these cases the D4FL-supported activities do not reach internally displaced children or refugees, unless they were in ordinary schools. Globally, progress on learning measurement models in EiEPC contexts has been disrupted by challenges to the aid architecture mandated to lead on it (see Annex 3, paragraph 28), which makes it more difficult for a programme like the D4FL, which is not focused on it, to shift the needle on reporting data for EiEPC contexts. Even so, there is no clear evidence of significant efforts – such as collaborating with NGOs or Education Cannot Wait (ECW), actions outlined in the D4FL Business Case – to provide more comprehensive information on children's learning in EiEPC contexts.³¹

²⁹ Paragraph 11 of the D4FL Business Case ([link](#)) outlines the rationale for work in EiEPC contexts, while paragraph 39 targets an increase in the number of countries reporting learning data in EiEPC contexts. Paragraphs 43–46 confirm that the programme will globally prioritise improved disaggregated data by sex, age, disability, and geography, and in EiEPC contexts work with non-governmental organisations and states to better track and support children's learning.

³⁰ The PAL Network is updating its tools to better include children with disabilities in the 2027/28 ICAN/ICAR round, but this work is not funded by the D4FL programme.

³¹ The PAL Network is working with ECW and Porticus to adapt the ICAN/ICAR tools for emergency settings, with the aim to embed this in future rounds of data collection. This work is not funded under the D4FL grant, however, and its inclusion in large-scale ICAN/ICAR rounds is planned for 2027/28 – after the D4FL programme has ended.

EQ2 To what extent do D4FL’s objectives and activities address the needs of the global community?

Finding 7. Most of D4FL’s programmed activities are aimed at filling the global foundational learning data gap, which is a need of the global education community.

87. In 2023 SDG 4.1.1a was demoted as an SDG indicator because of lack of data. This means that foundational learning and its measurement were to feature less prominently in the list of SDGs. A key factor in lack of data was lack of agreement at global level on which assessments would count for reporting. The D4FL focus on demonstrating ways to generate learning data that can meet global reporting requirements, so that more countries can report, should be seen in this light.

88. The learning data gap persists. The SDG 4 2025 scorecard shows that 68 percent of countries have not reported data on foundational learning, or have no trend data to measure progress towards SDG 4.1.1.³² In the September 2024 UIS data release, only 37 countries reported learning (using reading as a proxy) at the Grade 2 or 3 level, and only 101 countries have reported learning at the end-of-primary level at least once in the last six years. This is in sharp contrast to reporting of primary school enrolment where 203 countries report.³³

89. The D4FL programme supports more countries reporting in various ways. The D4FL’s support to UIS contributed to the agreement of a Global Proficiency Framework (GPF) and the publication of criteria to report on SDG 4.1.1a. The programme has also funded multiple service pathways for the generation of potentially compliant data, e.g. via integration into national assessments with technical support (ACER UK, World Bank ALMA), via AMPLs (ALMA, UIS), and via household surveys whether citizen-based (the PAL Network) or in conjunction with government (UNICEF MICS). Global respondents thought it helpful for countries that D4FL has resulted in a menu of potentially viable options.

4.2 Coherence

90. 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.

Summary of findings on coherence: Partners share a unified aim of improving learning outcomes and agree on the importance of quality foundational learning data to drive action towards this aim. However, incoherencies and disagreements within the D4FL partnership risk confusing advocacy efforts, and undermining the value of data produced through D4FL-funded activities outside of national government systems. Overall, D4FL demonstrates a high level of coherence with other global and national efforts to strengthen learning assessment systems, though this coherence is not yet fully consistent across all contexts. The programme is broadly seen as complementary rather than duplicative, aligning with existing national assessments and collaborating with key actors to support harmonised approaches to foundational learning data. In several countries, D4FL has added value by filling gaps in technical capacity, promoting alignment with SDG reporting requirements, and contributing to greater coherency. However, coherence is often dependent on strong local relationships and FCDO advisor facilitation, meaning coordination varies across countries. While global-level linkages are strengthening, room remains for more systematic joining-up of FCDO-funded programmes at country level and clearer pathways for how D4FL and related initiatives should complement one another.

³² UIS, 2025a

³³ UIS, 2025b

EQ3 To what extent is D4FL a coherent programme across strands and partners?

Finding 8. D4FL partners share the aim of improving learning outcomes and the objective of increasing the availability of (internationally comparable) foundational learning data, but there is a lack of coherence around what should be emphasised in reaching these objectives.

91. There is strong evidence that the D4FL partners share a common aim of improving foundational learning outcomes. Annex 4 shows how the work of each partner maps onto the programmatic theory of change, aiming to contribute to this ultimate impact. This aim was a) stated consistently by partners at country and global level as driving the work, and b) recognised by partners as a common goal that provided consistency and coherence across the programme.

92. Except for the Accelerator initiative, which shares the ultimate aim of improving foundational learning but focuses on strengthening education systems and practices, all other D4FL-funded activities share the objective of increasing the availability of quality data on foundational learning. As set out in Annex 4 this is consistently intended to lead to strengthened action on foundational learning, though through various pathways including national governments' direct use of data to inform curricula reforms and interventions, and in-country and global pressure for action.

93. Across partners – including ACER UK, PAL Network, UNICEF, the World Bank, and UIS – there is a consistent emphasis on generating reportable SDG 4.1.1 data while supporting countries to strengthen their own systems for learning assessment and data use. Multiple sources highlight global alignment around filling global learning data gaps, producing disaggregated and comparable data to inform national decision-making, and building capacity to use evidence to improve teaching and learning. This shared vision at the global level is reflected in joint efforts to align national assessments to UIS criteria, and in shared sector communications (e.g. conference panels) – some of which happened during the MTE and more of which are planned.

94. Despite these efforts to collaborate and build a shared narrative, there are also areas of incoherence. As described elsewhere in this report, the level of focus on international reporting varies across partners and across contexts. For instance, NFER's support to the EARC is primarily focused on capacity building and supporting government to strengthen the data use more broadly. While this might include data on foundational learning that can be used to report against SDG 4.1.1, this is not the primary focus. In contrast, the PAL Network's ICAN/ICAR initiative and World Bank- and UNICEF-led activity around AMPL and MICS are primarily focused on generating internationally comparable data and specifically data that can be used to report against SDG 4.1.1. TA provided by ACER UK and by the World Bank in some cases, meanwhile, seeks to strengthen national assessments to meet the UIS criteria. In these cases, the relative emphasis placed on international reporting vs. in-country use tends to vary between individual stakeholders.

95. Where TA is focused on international reporting, there appears to be a consistent view among implementing partners at the global level that aligning with the UIS criteria for SDG 4.1.1 reporting should serve to strengthen the availability of quality data on foundational learning in country, and that global reporting should not be seen as an end in itself or prioritised above in-country needs. However, as described in Finding 3 above, this is not always consistently playing out on the ground, and there are examples of global reporting needs being prioritised over strengthening government capacity and institutionalised assessments.

96. There is also a lack of consensus about the importance of working within government systems. The PAL Network model seeks to strengthen national capacity for learning data generation, but outside of government systems. Positive views of the strengths of this model among D4FL partners and stakeholders are that this is beneficial for sustainability, provides a model for data generation where capacity does not exist in government and provides complementary, independent data for accountability even where government data does exist. However, there is not agreement on this amongst D4FL partners, and some argue that working outside of government structures undermines long-term system strengthening. This lack of shared narrative risks confusing advocacy efforts, and risks the value of the data produced through activities other than TA to national government being undermined in the future.

Finding 9. Although partners share goals, coordination gaps risk incoherence or competing approaches to data generation.

97. These key risks are:

- fragmentation and parallel efforts
- institutional incentives that at times pull partners in different directions
- coordination and coherence relying on individual efforts

98. Fragmentation and parallel efforts can weaken coherence and could undermine the effective pursuit of shared aims. While D4FL partners broadly share objectives, coordination challenges at both global and country levels have occasionally led to overlap or uncertainty in roles. In several countries, notably Nepal, Ghana, and Somalia, similar assessment activities have been pursued by multiple partners without consistent mechanisms for joint planning or communication. In Ghana, at the time of our data collection in June 2025, the division of responsibilities for national assessments between the World Bank, ACER UK, and government agencies remained unclear, but this was reportedly resolved by end 2025. One survey participant from Rwanda highlighted that the UNESCO AMPL and ACER UK's work on national assessments did not overlap, but both "are requiring time of a very busy agency". This is a particular risk for demand-led models of support, where governments may be keen to access support and therefore risk inefficiencies or spreading themselves too thin across multiple partnerships. Although collaboration between partners over time has improved over time, strengthening joint planning processes and country-level governance arrangements could help mitigate duplication and improve synergy among partners.

99. Institutional incentives at times pull partners in different directions. Despite alignment on overall objectives, institutional mandates and proprietary models are sometimes perceived to drive fragmentation. Evidence from interviews indicates that organisations maintain strong attachments to their own tools and approaches, resulting in a crowded landscape of measurement initiatives. This dynamic, described by one respondent as an "oligopoly in measurement", can make collaboration more difficult and perpetuate inefficiencies. Similarly, partners may prioritise their institutional visibility or technical model over integrated system strengthening, particularly when operating in countries where multiple donors are active. These tendencies reflect underlying systemic challenges in the global assessment ecosystem, rather than D4FL design flaws *per se*, but they nonetheless constrain the programme's coherence.

100. Coordination and coherence, particularly at country-level, still rely heavily on ad hoc engagement and personal relationships rather than on structured coordination mechanisms or shared systems. While the D4FL partners have reported efforts to coordinate their work amongst

themselves and with governments (e.g. in Malawi where both the World Bank ALMA and ACER UK adjusted their work programmes in discussion with government), the coherence of D4FL's shared objectives is generally more visible at the global level than in-country implementation. At the global level, partners regularly participate in joint meetings, share reporting frameworks, and articulate a common commitment to foundational learning and data comparability. However, at the country level, evidence of joint activity is mixed. In contexts such as Kenya and Nepal, the PAL Network country partners worked in parallel without interaction with other D4FL partners, while in other settings, coordination depended on the initiative of FCDO education advisors. This reliance on individual initiative rather than institutionalised systems means alignment can be fragile, particularly when staff change or funding cycles shift.

EQ4 To what extent are D4FL activities coherent and complementary with other global learning data efforts?

Finding 10. The evaluation found no evidence of duplication of other efforts, and there are some strong examples of D4FL work being tailored to complement and support other initiatives.

101. Evidence from global and country sources indicates that D4FL activities are largely coherent with, and complementary to, other global learning data initiatives. Across contexts, stakeholders reported no duplication of existing efforts and, in several cases, highlighted that D4FL fills important gaps in the learning data landscape rather than overlapping with established mechanisms.

102. The World Bank's ALMA initiative provides a strong example of complementarity, working alongside regional assessment programmes such as PASEC and SACMEQ to analyse existing data and address implementation gaps where regional assessments have partial or incomplete coverage in countries. Similarly, the PAL Network and UNICEF are exploring aligning MICS and ICAN/ICAR.

103. ACER UK's support to countries has focused on strengthening and aligning national assessments with UIS criteria for SDG 4.1.1 reporting, thereby enhancing rather than replacing existing systems. Country evidence from Ghana, Kenya, and Somalia confirms that D4FL-supported work is additive, with stakeholders perceiving it as tailored, responsive support that builds on what already exists rather than creating parallel structures.

Finding 11. D4FL has also contributed to greater harmonisation of global learning assessment approaches, supporting dialogue and helping align standards across implementing partners and global actors.

104. We found that UIS, FCDO and key actors such as the World Bank are increasingly working together to articulate a common narrative on foundational learning data, including through the Global Coalition for Foundational Learning and joint technical exchanges. Evaluation participants reflected that FCDO's convening role has helped strengthen linkages across donors, technical agencies, and implementing organisations, contributing to a more unified global narrative around foundational learning measurement. Interviewees from UIS, the World Bank, and country governments noted that D4FL has played a stabilising and enabling role, particularly by reinforcing alignment with SDG 4.1.1 and promoting coordination rather than competition among actors.

105. As discussed in Finding 3, the D4FL programme has for the most part supported countries to navigate global requirements in ways that reinforce rather than displace national systems. The

programme has also supported collaborative efforts, such as the PAL Network's ICAN/ICAR initiative, which is recognised as bringing coherence to previously fragmented national efforts and enabling technically sound, internationally comparable measures that retain relevance in-country.

106. However, while no direct examples of duplication were identified, some limitations in coordination and harmonisation remain. Several participants observed that global coherence across the learning assessment ecosystem is still evolving, with persistent fragmentation and a proliferation of tools posing risks to alignment (see paragraph 99 in Finding 9).³⁴

Finding 12. Despite progress in aligning global actors around foundational learning measurement, incoherencies in the international assessment landscape persist and play out within the D4FL programme and coalition.

107. Evidence indicates a lack of consensus among programme partners on whether generating and tracking internationally comparable data on foundational learning should be a globally-driven or country-driven priority. While some global actors perceived providing TA for global reporting against SDG 4.1.1 as an ambition for all nations, several perceived this as an unhelpful, "top-down" dynamic. One view was that national governments must drive global reporting, and therefore that TA to achieve SDG 4.1.1 reporting should not be funded if the desire for global reporting is not reflected in a country's national education sector plan. Another view was that internationally comparable data can be driven by national or regional stakeholders other than government.

108. In some contexts, governments face pressure to generate data that meet SDG 4.1.1 or UIS eligibility criteria, even where this may not align with national policy cycles, system capacity or the most pressing evidence needs for improving teaching and learning. Interviewees noted that global reporting demands can result in assessment designs that privilege comparability over usability, leading to data that are less relevant for classroom-level decision-making or national reform.

109. In addition, we found that government buy-in to the SDG 4.1.1 reporting criteria is emerging but not yet universal. National governments that have received tailored technical support show stronger alignment and growing ownership of the criteria, with Ghana as a leading example (see Finding 3 above and Finding 18 below). However, in several contexts the criteria are still viewed as externally driven or technically demanding, and consensus on their application remains incomplete.

110. Finally, stakeholders noted that there is still no unified consensus on the requirements for global reporting, with the UIS vetting mechanism and SDG 4.1.1 eligibility criteria continuing to evolve (an agency was appointed only in October 2025 to pilot a vetting approach, and the criteria are not yet formally approved). While the UIS "middle path" approach seeks to balance standardisation with flexibility, there remains ambiguity around how much alignment and what level of technical rigour is expected of countries. Actors did not have a shared view on how the newly developed UIS criteria would be applied, with some actors hoping the criteria will be used as aspirational targets to encourage progress in assessment quality, rather than a "pass" and "fail". There were also particular criteria that assessment actors were unclear about – for instance whether the requirement for a field trial to be "representative" will be interpreted to refer to statistical representativeness or a more qualitative criterion (i.e. the diversity of schools and learners to be included in a sample-based assessment are included in the trial).

³⁴ With this said, the evaluation team notes that reducing the available global tools may align assessments at the global level, a reduction in tools would significantly impair efforts to align with national systems.

EQ5 To what extent is D4FL coherent and complementary with other FCDO programmes?

Finding 13. Complementarity is intended by design and has potential for impact.

111. The evaluation found that D4FL and other FCDO CMPs are designed to be complementary. Through the Learning at Scale pathway, evidence generation (D4FL), evidence use and decision-making support (What Works Hub for Global Education (WWHGE), Scaling Access and Learning in Education (SCALE)), and delivery/implementation (bilateral programmes) are intended to reinforce one another along a common reform trajectory.

112. When coordination is effective, this model has strong potential: in Somalia, for example, the FCDO advisor facilitated links between ACER UK's assessment work and a bilateral Teaching at the Right Level (TaRL) initiative, ensuring that assessment data will inform teacher training. In Kenya, recent progress – such as NFER's work with KNEC starting to join up with What Works Hub activities – indicates movement towards cross-programme alignment.

Finding 14. Complementarity is driven primarily through FCDO Education Advisor actions and collaboration at global level, and is inconsistent.

113. Despite the strong theoretical links between CMPs and bilateral programmes, evidence of this complementarity translating into practice in countries remains limited and inconsistent.

114. Because programmes are demand-led, linkages across CMPs occur only when in-country actors recognise the opportunity and initiate the connection, resulting in uneven implementation of the Learning at Scale pathway and limited integration into broader decision-making channels. Coordination is largely informal and relies on individual leadership from proactive FCDO advisors or well-connected programme partners, rather than institutionalised mechanisms.

115. Links with other FCDO CMPs were not spontaneously mentioned by any interviewees across our country cases, though the latest progress report from NFER indicates plans to link up with the work of WWHGE. Where advisors were absent or not engaged, as seen in Nepal, Kenya and Mexico, some partners were unaware of other FCDO programmes, and no active links were being made. Even in Ghana, where the FCDO advisor has played a strong convening role, we observed that the EdTech Hub's work on data dashboards did not come up in interviews with the World Bank and UNICEF partners who are working in the same space. In the survey, participants from Sierra Leone and Rwanda specifically highlighted that better links could be made between the D4FL programme and WWHGE.

116. Despite these challenges, stakeholders generally perceived CMPs as additive rather than overlapping. There is also growing momentum towards cross-programme learning, with shared partners (e.g., Innovations for Poverty Action (IPA), British Council) and FCDO-supported platforms helping align methods and reinforce coherence across initiatives.

Finding 15. Strengthening the links between D4FL and other CMPs and bilateral programming could increase the likelihood of achieving D4FL's *Use* outcomes and ultimate intended impact of improved foundational learning.

117. Stakeholders noted that strengthening coordination could help reinforce the *Use* strand of the D4FL programme, by connecting assessment data not only to technical experts but also to policy, teacher professional development, and system reform actors. To realise the full potential of

the Learning at Scale model, stakeholders recommended institutionalising country-level coordination mechanisms such as joint workplans, routine cross-initiative meetings and clearer architecture, to reduce reliance on individuals and ensure that coherence is sustained through turnover and context change.

4.3 Effectiveness

118. Effectiveness refers to the extent to which D4FL is likely to achieve its intended outcomes.

Summary of findings on effectiveness: Available evidence indicates that D4FL can deliver on its objective of filling global learning data gaps, through support for the generation of foundational learning data for 27 countries that can be submitted to UIS for vetting. In delivering this support, partners have used methodologies that can be replicated to track foundational learning elsewhere. There is, however, still uncertainty about whether the data will be accepted in all instances, and it has made little progress on filling gaps for children affected by emergencies and protracted crises. The D4FL programme is likely to deliver on its objective of supporting more or better-quality disaggregated data that can be used to design and monitor foundational learning interventions, as D4FL's support strengthens assessments and assessment capacities in countries. In a limited number of countries, D4FL-funded activities are likely to lead to data that are more easily accessible for use, or made useful for application and decision-making in country. D4FL has not, however, strengthened capabilities to use data in any significant way. Its contribution to filling country gaps on learning data for populations that are often excluded, is also likely to be limited.

EQ6 To what extent is D4FL leading to the availability of more disaggregated learning data in LICs and MICs that can be used to design and monitor foundational learning interventions?

Finding 16. D4FL is close to delivering more and more comparable or better-quality foundational learning data in 19 supported LICs/LMICs, and processes are under way to generate more or better data in more countries.

119. D4FL is supporting foundational learning assessments in 24 countries, two regions and UNRWA through the World Bank Alma and Accelerator, ACER UK, UNICEF, NFER, and PAL Network activities (see Figure 1 above for an overview and Annex 5 for details). In seven of these, two or more partners are undertaking the work. At the time of writing, the evaluation evidence indicated that better quality or more comparable foundational learning data will likely be available in 16 of these 24 instances of support. In Ghana, Rwanda, the Caribbean regional intervention (four countries), and for the PAL Network 13 countries, significant progress has been made, and data were on the cusp of being reported at the time of the evaluation's data collection. In the other eight instances, work was in too early a stage (agreeing scope of work, inception, contracting and delivery of TA) to conclude likely success.

120. The evaluation survey results provide some confidence in the likelihood that D4FL-funded interventions implemented by World Bank ALMA and ACER UK will increase the availability, quality or comparability of foundational learning data. Most of the respondents from the countries where these two partners worked were aware of the work and thought there would be benefits and no negative consequences from the work (see Annex 9 Table 21 and Table 23 for more detail). The survey provided less certainty about the outcomes of the PAL Network assessments, reflecting low knowledge about what the Network was implementing in the D4FL-supported round of ICAN/ICAR

assessments, as well as a perceived longer and sometimes challenging route from the Network’s ICAN/ICAR assessments to the use of data for improved learning (see Annex 9 Table 22).³⁵

121. The box below provides an overview of D4FL investments and the extent to which they will contribute to data coverage and quality improvements. It shows that although the programme is focused on supporting the generation of data that can be used to report against SDG 4.1.1, in at least 14 countries, it is also supporting more or improved data for national use.

Box 2 D4FL contribution to improvements in learning data coverage and quality

More data in 3 countries: In **Somalia** there is no nationally representative data on learning before the end of primary school. An ALMA AMPL will deliver this data, and ACER UK’s development of a national assessment framework helped government target such data through national assessments. In **Nepal**, where there is a lack of foundational learning data³⁶, the D4FL is supporting an AMPL through ALMA and the PAL Network ICAN/ICAR assessment, which will deliver this data. In **Senegal** Accelerator is supporting local language reading assessments in two regions, which previously had no data.

More coherent data in 2 countries: In **Sierra Leone** (*not a case study*)³⁷ ACER UK will review the many existing assessments for UIS integration and build capacity of the national assessment authority, established in 2021, for assessment design towards regular, comparable and coherent assessments. In **Malawi** (*not a case study*), early support by D4FL had a similar objective. Subsequent support is to conduct a review of localised assessments to determine a single assessment to support resource allocation.

More accurate, relevant, temporally comparable data in 8 countries: D4FL supports more temporally comparable assessments and data. For example, in **Ghana**, ACER UK worked on more temporally comparable data through test design during its support to integrate the UIS criteria into national assessments, and ALMA is working in **São Tomé and Príncipe** (*not a case study*), comparing assessments through support for analysis. The D4FL programme supports the improvement of assessment design in **Rwanda** (*not a country case*), where ACER UK supports on contract management of outsourced assessments; in **Djibouti** and **Mali** (*not country cases*), where ALMA is building capacity on test design principles and the construction of an item bank; in Djibouti, also on sampling and sampling plans; in the **Kyrgyz Republic** (*not a country case*), on large-scale test design, implementation and reporting; in **Lao PDR** (*not a country case*), for the development of Grade 3 National Assessment Framework, aligned with the recently approved curriculum, and test specifications linked to the GPF; and in **Somalia**, ALMA is reviewing country capacity for assessment to identify capacity building needs.

More (and better) inclusive data: The PAL and MICS assessments (**14 countries, including four country cases**), provide nationally representative data and are household assessments, thus generating learning data that are more inclusive of populations often excluded from school-based assessment (OOSC, children in rural areas, children in private and community schools). The D4FL programme will contribute to more gender-disaggregated data where countries supported to report on SDG4.1.1. are not currently disaggregating by gender (unclear how many countries). The **UNRWA** (*not a case*) support applies global good practice principles in an EiEPC context to strengthen the quality of learning data for Palestinian refugee children. Where countries were previously assessing often-excluded populations, D4FL’s support may improve the quality of this data as a corollary of improving the quality of the assessments.

Source: Country case study and global KIIs, partner workshops, partner D4FL reports

³⁵ The survey did not test country stakeholder perspectives on NFER and UNICEF, as D4FL support only started in 2025.

³⁶ Multiple evaluation participants reported this lack of foundational learning data, though we are aware of a [National Assessment of Reading and Numeracy](#) in Grade 3, carried out in 2020.

³⁷ Countries that are not country cases are noted. In these cases, data are not triangulated through country-level research.

Finding 17. D4FL is likely to lead to data that is available for use and made useful for application and decision-making in country in some, but not in all supported countries.

122. The evaluation has not been able to observe directly whether and how D4FL support is contributing to data that can be used for in-country decision-making, mainly because the support had not yet led to data being released/reported by end September 2025. The team therefore looked for evidence of whether support is likely to generate data that are available for use, and useful for application and decision-making. In this regard we looked at evidence of the intent to make data accessible, to align data with country needs, to analyse data so that it can be used (and build country capacity for this analysis), and examined past evidence of usefulness of data (or not) that might be replicated for the D4FL-supported assessments.

123. In total, D4FL-supported interventions include specific actions directly with country authorities that make use more likely in 13 out of the 24 countries, regions and authorities supported for data generation. The composition is set out in Table 3 below, drawing on D4FL partner reports, with triangulated examples from the country cases.

124. It is worth noting that these are merely a count of possible cases where D4FL has activities that make these links: we have not yet seen the data reports, dashboards, analysis and support outcomes to assess the quality of these interventions and their likelihood of success.

Table 3 Instances of D4FL support that build links from data generation to use

Type of support towards use	Instances	Explanatory examples from the country cases
Working to establish use as a priority assessment objective	Somalia, Malawi	In Somalia, ACER UK’s support for the development of a national assessment framework, led to the framework prioritising assessment tools for classroom use by teachers.
Making data more accessible for data users	Ghana, Somalia, Rwanda, UNRWA	Support for setting up an assessment data dashboard in Ghana (ALMA); linking Somalia AMPL data and results to the data and results of the Global Education Policy Dashboard (ALMA).
Support to ensure that data and analysis address specific learning needs and data gaps, and can respond to requests from policy makers	Djibouti Guyana Lao PDR Malawi, Mali Rwanda São Tomé and Príncipe Senegal UNRWA NFER/EARC	In Senegal, the World Bank Accelerator is conducting a local-language EGRA to monitor early grade teaching reforms. In the NFER/EARC support, the focus is on assessment capacity to respond to national use needs, and train on assessment data analysis for use “We focus on data for decision-making and making the right choice of assessment from the start rather than trying to learn through very expensive lessons because of choosing the wrong assessment or implementing the wrong way.” In Ghana, D4FL (ALMA and ACER UK) is supporting capacity building for the analysis of assessment data. In Malawi, the ALMA programme is planning to support analysis of SACMEQ V data, as well as dissemination of the analysis to support decision making.

Source: D4FL partner reports, global level interviews and workshops with partners, and country case study KIIs and document review.

125. D4FL’s support for the PAL Network could lead to the use of ICAN/ICAR data in education policy decisions, programme designs and interventions at different levels of the system, even if data

were not generated by the authorities. Whether such application of the data will happen, will be context-dependent and remains to be seen for this D4FL-supported round of ICAN/ICAR.

126. However, there are three factors that support a higher likelihood that it would do so. Firstly, country authorities were involved in the sampling frames. Secondly, the Network has a D4FL-funded advocacy strategy in place that outlines which target partners, allies, and potential opponents they will work with and how they will ensure the results of the assessments inform decision-making and action, an example of planning for use before the data arrives. Thirdly, in addition to the evidence from an earlier evaluation on how PAL Network assessments have influenced national discourses and policy commitments (see paragraph 79 above), the Mexico country case study provided evidence of how the PAL Network partner there has worked in the past to ensure data use (see Box 3 below).

Box 3 The case of Mexico: building bridges to use of citizen assessment data

In the past, PAL Network partner in Mexico, Medición Independiente de Aprendizajes (MIA), has generated formative classroom assessment data that fill the diagnostic gap left by the suspension of national assessments in Mexico. While the same dissemination, training, and advocacy actions would not be possible from nationally representative ICAN/ICAR assessments, it serves to illustrate the effort that PAL Network members put into post-assessment advocacy and follow-up. Past tests provided data on reading, math, and socioemotional skills, aligning with local government and civil society priorities. Results were easy to interpret, presented in visual, school-level reports that helped teachers and principals act quickly. Training was also provided to educators, amongst other, to use and interpret that data. It combined in-person workshops, online learning modules, and cascade models that reached schools through trained intermediaries. Educators were taught how to interpret results, identify learning gaps, and implement remedial actions. The data have been used to design school improvement plans, guide state recovery programmes, and even inform the 2022 National Educational Trajectories Strategy, which reached more than 80,000 teachers. Although not formally adopted by federal authorities, data generated by MIA are regularly used by NGOs, donors, and local education actors to guide resource allocation and to advocate for stronger foundational learning policies.

Source: Mexico country case study done for the evaluation, June 2025

127. The evaluation has no evidence yet that global reporting of comparable data based on the UIS criteria, creates pressure for country action on learning in D4FL countries and beyond, as is assumed in the D4FL theory of change. Country case respondents as well as evaluation survey respondents have pointed out both the potential benefits of global and regional comparisons of learning outcomes, as well as the political risks for continued testing if results are negative.

128. These views align with evidence from past studies on the impacts of international and regional assessments like PISA, SACMEQ and PASEC.³⁸ These studies have shown that positive impacts from assessments reform occur when contextual information about teaching and learning conditions is collected, so that the results can be used to design education interventions. Other conditions for impact include that the tests are credible and there are strong dissemination of results and engagement with stakeholders. There are, however, evidence of negative consequences, e.g. the results become political scoring points preventing a serious look at policy and inappropriate policy borrowing from countries that fare better.

³⁸ As set out in Cambridge Education, 2020, SEA-PLM Literature Review, Unpublished.

EQ7 To what extent is D4FL leading to better capability in priority countries to use learning data to design and monitor foundational learning interventions?

Finding 18. Where D4FL supports generation of learning data in national assessments, country governments have clear ownership of the data.

129. The evaluation has evidence that D4FL is leading to assessments and learning data that are owned by decision-makers, even if in some cases ownership of assessments and data predates the D4FL interventions, such as in Ghana. Evidence includes:

- *Consultative or participatory processes:* An early indicator of potential country ownership is the extent to which government actors participate in processes to design and implement the support. This was observed in three of the country cases, Ghana (joint hands-on workshops on test design and for pairwise comparisons), Somalia (participation in consultation for the draft framework) and the NFER support to KNEC for the EARC in Kenya (co-development of implementation plans for KNEC).
- *Country authorities taking on a higher burden in subsequent rounds, with D4FL assistance for assessments taking a back seat:* In Ghana, the assessment authority has taken the lead in item writing for the 2025 P2 national standardised test, following hands-on workshops with ACER UK to review items for the 2024 P4 assessments. In Somalia, the education ministry ran assessment framework consultation processes. In the case of the NFER support for the KNEC, KNEC took a strong role in the EARC launch workshop with regional representatives.

130. We also have evidence that ownership is not as evident in the two country cases where an ALMA is being implemented, despite efforts to build ownership. In Somalia, the World Bank reported that there was wide consultation with the education ministry on conducting an AMPL separately from the planning of a first national assessment. However, education authority respondents did not see an AMPL as a priority relative to the first national assessment. In Nepal, government stakeholder engagement has been inconsistent. The World Bank team has tried to get the Education Review Office to take ownership by undertaking workshops and tried to build appetite for the use of the data. However, at the time of the country case study, lack of government engagement indicated that implementation would be fully parallel to the national assessments.

131. The contrast in ownership between country cases where national assessment systems are adjusted for global reporting, and cases where an AMPL is being done, underscores that achieving the dual goals of global reporting and better national data generation for country use simultaneously may be more difficult when there is no ownership of the assessments, even when D4FL implementing partners try to engage officials in processes to administer the assessment.

Finding 19. There is evidence in only two country cases that D4FL is building capacity for data use beyond assessment authorities.

132. Finding 17, in Table 3 above, presented cases where D4FL is supporting country assessment units to undertake analysis. In addition, there are two cases of D4FL support for building the capacity of broader education actors to use assessment data reported in D4FL partner reports. In Guyana, ALMA's scope of work includes TA and capacity building to use assessment results for remediation programs; and in Lao PDR, ALMA is building the capacity of policymakers and leaders in the education ministry to use learning assessment results for policy and programming. The evaluation has not triangulated the effectiveness of this support with country stakeholders.

Finding 20. Country cases suggest D4FL-generated data will be used in foundational learning policies, targets and interventions, but only in a few cases.

133. Data supported by D4FL will or is likely to be used by countries primarily, but also donors in developing and monitoring policies, targets, and interventions in eight of the 24 countries, regions and authorities supported by D4FL, as analysed in Table 4 below. This use may be passive (use already established prior to D4FL or not supported by it) or active (use supported by D4FL).

Table 4 Country cases where D4FL support is likely to lead to use of data

Type of evidence	Examples (<i>Note, non-case study instances draw on partner reports and interviews; this evidence is not triangulated</i>)
Cases where D4FL directly supports the use of D4FL-supported data	<ul style="list-style-type: none"> • In São Tomé and Príncipe (<i>not a case study</i>), ALMA supports the drafting of a costed action plan for responding to findings of learning assessment. • In Mali (<i>not a case study</i>) the analysis done with D4FL support is expected to be used to guide curriculum reform and teacher training.
Cases where past use of data signals likelihood that D4FL-supported data will be used	<ul style="list-style-type: none"> • In Ghana, the 2022 and 2024 P4 and P6 assessments will be used to evaluate the Ghana Education Outcomes Project, aimed at reintegrating OOSC into formal schooling (the NSTs are census-based assessments). The National Teacher Association, a teachers’ union, also uses the data to design its teacher professional development interventions. • In Mexico, past use of PAL Network data, is promising for use of the D4FL-supported 2025 ICAN/ICAR round (see Box 3 above).
Cases where there is clear future intended use of data	<ul style="list-style-type: none"> • In Senegal, the local-language EGRA supported by Accelerator will be used to monitor the government foundational learning programme. • In Ghana, the Ghana Accountability for Learning Outcomes Programme (GALOP) will utilise NSTs at district and school level, to monitor the effectiveness of interventions in the targeted low-performing schools, including by requiring that data for the school is posted publicly. • In Somalia, the intent was to analyse data from the first national assessment to guide teacher professional development through multi-donor contributions. • In Rwanda (<i>not a case study</i>), Learning Assessment in Rwandan Schools (LARS) data is being used by Tetra-Tech and Innovations for Policy Action for baseline evaluations. These evaluations are linked to donor-funded programmes, creating external pressure for measurable improvements.
Cases where there is a significant up-front effort to ensure that data will respond to country needs	<ul style="list-style-type: none"> • In UNRWA (<i>not a case study</i>), ACER UK has set out a programme of support that includes significant consultation with stakeholders, including teachers, to understand needs from assessment. This support can translate into actual use. • In Malawi (<i>not a case study</i>), ACER UK and the World Bank ALMA have worked together to assess what country needs are before initiating work on assessments. Outcomes include support to review localised assessments and to design a common assessment to guide national, regional and local resource allocation.

Source: Country cases, global interviews and workshops with partners, partner D4FL reports.

134. The two country cases that are not detailed in the table, Nepal and Kenya, offer insight on the barriers D4FL can face to translate its support for data generation into data use for country action. In Nepal, where poor use of data was already a challenge according to country KIIs, the lack of ownership discussed in Finding 18 does not support likely use of the AMPL data. In Kenya, where D4FL supports the PAL Network to generate citizen data on Kenya’s learning outcomes, the PAL

Network partner perceived the exercise to primarily be about regional and global comparability, and – at the time of the MTE data collection – reported no plans to advocate for its use at national level. That said, global representatives of the PAL Network describe a phased approach, where 2025 activities were intentionally focused on global reporting while national reporting – including equipping country teams to produce and disseminate country-specific reports – is a priority for 2026.

EQ8 To what extent and how is D4FL filling global learning data gaps?

Finding 21. While no countries supported by D4FL have reported and been vetted yet, D4FL will likely result in 27 countries having comparable data complying with UIS criteria; out of these only two are reasonably certain to submit data for vetting soon.

135. Through ACER UK/PAL Network/UNICEF support, assessment data for 27 countries integrating the UIS criteria could be available by the end of the programme. Of these, eight countries have reported before. However, for all D4FL-supported countries, not all data will necessarily be used to report on SDG 4.1.1 to the UIS. There are uncertainties still about whether all the assessments will be successfully completed, whether supported countries will be willing to submit the data, and whether the UIS vetting process will accept the data.

136. Given current progress in the implementation of D4FL-supported interventions, we can say with reasonable certainty that data for two countries will be submitted to the UIS vetting process, and if accepted can be used to report against SDG 4.1.1. These are Ghana and Rwanda, neither of which have reported data before. In eleven further cases, if D4FL's support concludes successfully, countries are likely to submit data for vetting.

137. In six countries, reporting is more uncertain. Other are (i) two PAL Network (only) countries, where it is not confirmed that they will be able to report data to UIS even if ICAN and ICAR are compliant with the criteria (see Box 4 below); and (ii) the four countries of the East Caribbean Organisation of States supported by ALMA where data is expected soon, but it is not clear whether the responsible governments will submit the data for vetting.

Box 4 Support for the PAL Network and closing the gap on SDG 4.1.1a reporting

Members of the PAL Network have raised questions about whether ICAN/ICAR data can be submitted for global reporting by national actors other than ministries of education. However, UIS representatives reported to the evaluation team that they had clarified with the PAL Network and others that it would be the prerogative of governments to decide which assessments to use for global SDG4 reporting. This means that the PAL Network – and potentially other partners such as UNICEF – will have to overcome the hurdle of government accepting their data and using it for reporting. The feasibility of this will likely vary between contexts, depending on the relationship between the civil society membership organisation and government, as well as the extent to which their results are compatible with the government's narrative. That said, initiatives like ICAN/ICAR and MICS can provide internationally comparable data and findings, even if not used for SDG reporting.

Source: D4FL MTE interviews and written comments on the draft report

Finding 22. D4FL will likely contribute to establishing and piloting recognised methodologies that can track foundational learning, but not necessarily for all.

138. D4FL has contributed to defining what MPLs are for foundational learning and the establishment of criteria of minimum standards for assessments through its support to UIS. These

two interventions alone already should help countries report. Evidence, however, is mixed on whether the criteria set the bar too high, as discussed in Finding 12 above, and Finding 37 below.

New methodologies to track foundational learning

139. In addition, compelling evidence is emerging that by 2027 D4FL will have contributed to several assessment methodologies that countries can use to report on SDG 4.1.1. These are:

- The integration of UIS criteria into national assessments (implemented by ACER UK and the World Bank ALMA). ACER UK is planning to make the tools it has used available for other actors to apply, and establish a community of practice to support peer learning. This approach is likely to need highly skilled and relatively scarce TA to implement.
- The AMPL, first applied in 2021, an existing replicable global assessment methodology, is seen as having a low participation cost for government but requires UIS to agree to its use and translation, and for countries to provide in-kind contributions of staff time, space, etc. Although this appears lighter touch than aligning national assessments, these processes can be challenging (e.g. in Nepal) and without built-in co-benefits of directly improving national assessment for use.
- D4FL supported the integration of UIS criteria into the UNICEF MICS Foundational Learning Module (FLM) 2.0, that can be replicated across countries drawing on UNICEF capacity and support. UNICEF's FLM 2.0 and harmonisation work also signal momentum toward standardised yet adaptable approaches that could feed into global reporting.
- D4FL has supported the PAL Network ICAN/ICAR integration of UIS criteria. The ICAN/ICAR approach is also a standardised, but adaptable approach. Countries could use ICAN/ICAR data for reporting, where governments agree to do so. Even if countries do not recognise the data for SDG 4.1.1 reporting, the data would still likely be globally comparable.

140. These efforts collectively represent a shift toward multiple credible methods rather than a single model, and D4FL appears to be on track to provide "proof of concept" for at least some of these methodologies. The ICAN/ICAR tools are now understood to meet UIS standards (subject to the vetting of national documentation and data), marking a significant milestone in creating a credible, low-cost, and context-sensitive tool for LICs and LMICs. Work by ACER UK and ALMA (particularly in Ghana as the most advanced example) is likely to demonstrate that it is feasible to integrate UIS criteria into a national system using targeted TA.

141. There is growing recognition of the value of these approaches. UIS, GPE and key donors have engaged with D4FL-supported methodologies and are beginning to integrate them into guidance, advocacy, and shared learning platforms. Stakeholders also noted that D4FL has helped elevate the legitimacy of "good-enough, credible" approaches for LICs/LMICs – an important shift in the global assessment community, which historically has favoured more resource-intensive models.

142. However, at this mid-term stage, no D4FL-funded data has yet been vetted and accepted for reporting. While implementing partners anticipate success, two risks remain for D4FL delivering 'proof of concept' on the methods supported: how the vetting process will use the criteria, and whether governments will endorse data generated through non-state activity. It is not clear what a negative vetting outcome would mean for proof-of-concept of multiple credible methods for countries to report on SDG 4.1.1. It is likely to first trigger global engagement on where the bar is set and whether it is reasonable. The reasons for rejection will also be telling for the validity of different models to count as globally comparable data, even if not accepted for reporting.

143. Moreover, even with proof of concept for countries supported, some D4FL-supported methods still rely on external TA, specialist software, and psychometric expertise which may challenge how easily approaches can be replicated elsewhere. While the D4FL programme has supported country teams to work through the requirements, stakeholders emphasised that the technical bar remains high. Meeting the criteria typically requires advanced psychometrics, data quality assurance processes, and documentation that many countries are only starting to build.

Inclusive assessment models

144. As is noted in Finding 6, the D4FL programme is unlikely to directly contribute to delivery of models to include children in emergency schooling, children with disabilities and other often excluded groups in global learning data, other than on including children that would not be assessed in school-based assessments, through household surveys (the PAL Network, MICS).

Global public goods

145. The D4FL programme is likely to lead to the production of global public goods on assessment. Examples are the UNICEF Learning & Skills Data Lab that will make foundational learning data and analysis available; the ICAN/ICAR data and assessment tools; and the World Bank Accelerator analytical education intervention products. In addition, NFER is considering making available the tools and materials they are providing to the KNEC for setting up the EARC; and ACER UK is planning to make its tools available online, for others to replicate.

4.4 Efficiency

146. Efficiency refers to the extent to which the D4FL programme is set up and making best use of its resources to maximise the chance that programme outcomes will be achieved. Under this criterion we look both at allocative efficiency (EQ9) and operational efficiency (EQ10).

Summary of findings on efficiency: Broadly, D4FL's allocation of its resources, both by FCDO and partners, is in line with its aim and objectives. Most of its resources are ostensibly allocated to the *Track* strand of work, which could raise concern about its cost-effectiveness in delivering on addressing global learning poverty, given the absence of significant evidence on an effective link from global SDG 4.1.1 reporting to country interventions to strengthen learning outcomes. However, because D4FL-supported interventions in most cases have opted to adapt existing national assessments for global reporting, these interventions also contribute to strengthened country assessments, with links to use of data to develop and monitor education policies and interventions in many. This is an encouraging signal on cost-effectiveness, strengthened by the programme crowding in and leveraging resources from other sources towards its aim. A potential detracting factor, however, is the extent to which replicating D4FL-supported models would depend on the availability and cost of what is seen as relatively scarce global TA. Although programme implementation has faced delays, these are not concerning as they are being managed and mitigated.

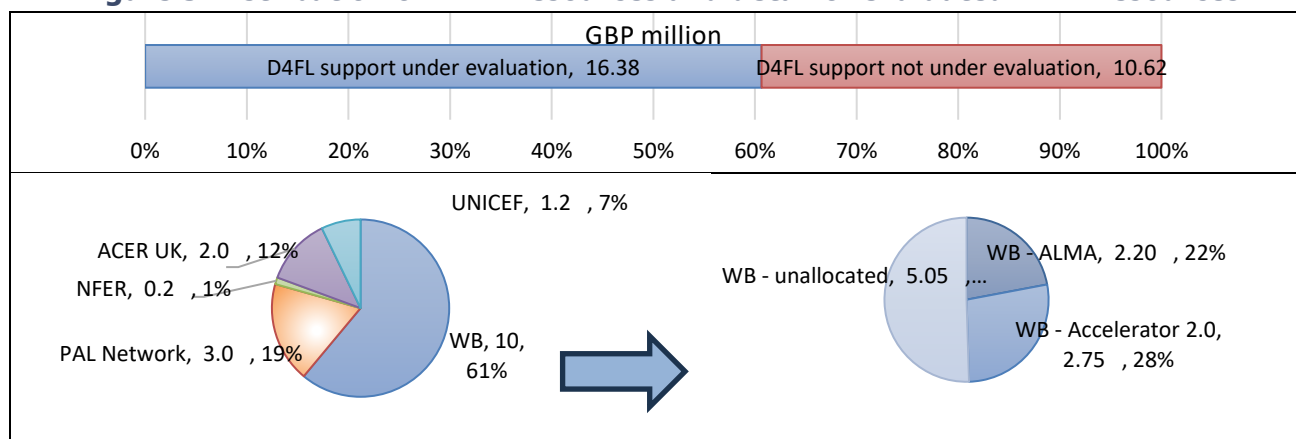
EQ9 How well is D4FL allocating resources to maximise the chance that outcomes are achieved?

Finding 23. D4FL has predominantly allocated funding to partner programmes that are in line with its objectives.

147. Most activities evaluated contribute to the programme's outcome objectives, as evidenced through the partner ToCs and through the review of active D4FL-supported interventions done for the MTE (see also findings under 4.1 Relevance).

148. The D4FL programme has allocated money to Accelerator 2.0 (17 percent of all resources under evaluation). As discussed in the formative evaluation, Accelerator 2.0 contributes to the overall impact, but its design does not support any of the D4FL outcome objectives (*track, generation or use*) directly. In practice, country cases and global interviews have revealed three cases where Accelerator 2.0 activities contribute to generating data or to increasing demand for the use of data. In Senegal, the Accelerator grant has been used to support Government in undertaking a learning assessment in three regions using EGRA, and the World Bank team noted that supported in Cambodia and Rwanda generating demand for reliable learning data.

Figure 3 Distribution of D4FL resources and detail of evaluated D4FL resources



Source: D4FL contracts and reports. Please note that unallocated in the right-hand side graph refers to the difference between the financing agreement commitments, and the amounts disbursed by the FCDO to the ALMA and Accelerator programmes by October 2025.

Finding 24. Investing in learning assessments is considered an efficient way of contributing to education outcomes.

149. Research shows that providing information on the benefits, costs and quality of education is one of the most cost-effective education interventions.³⁹ D4FL programme objectives make a direct contribution to this area by supporting the generation, tracking and use of foundational learning data essential to understand the quality of education and inform decisions to improve it.

Finding 25. The D4FL programme’s investment in data generation for global reporting through the alignment of national systems to the UIS criteria is an efficient use of its resources because it is benefitting the reliability and quality of data for country use.

150. The programmatic ToC (see Annex 4 Figure 6) sets different pathways in which D4FL investments can “contribute to progress against SDG 4.1.1 and learning poverty reduction in developing countries”. Identifying those that are most effective and understanding the optimal balance across them has implications for ensuring the efficiency of D4FL investments.

151. The extent to which D4FL partners align with or contribute to the three strands (*Track, Generate and Use*) is relevant from an efficiency perspective. In the D4FL ToC it is possible to observe a first direct route which focuses on data generation, tracking, and use at country level. A second route sees the generation of global and comparable data as a way to increase demand for data and education interventions at country level.

³⁹ GEEAP, 2023. This may change as development resources shrink, and funding sources shift to different sets of decision makers.

152. There is limited evidence that the act of reporting globally comparable data on SDG 4.1.1 on its own will stimulate demand for more data at country level (see Finding 38 below), or lead to action on learning (Finding 17 above).

153. However, investing in data generation for country use, versus data for global reporting is not a zero-sum game, as alignment of assessments to criteria for global reporting can also benefit the reliability and quality of data for national use. As was emphasised by D4FL partners, the UIS' SDG 4.1.1 reporting criteria⁴⁰ set minimum standards for learning assessments for global reporting that can also serve the broader purpose of better-quality national assessments for use. Such a set of standards that are formally agreed by a mandated global body with country representation – such as through the Global Alliance to Monitor Learning⁴¹ and the Education Data and Statistics Commission' (EDS Commission) – have been absent in the field.

154. Most of D4FL's support to country authorities for global tracking is indeed to integrate the UIS criteria into national assessments (ten countries and one region supported by ACER UK and/or the World Bank ALMA programme).⁴² As a result of these investments, as is discussed in Finding 3 and Finding 16 to Finding 18, while the D4FL programme is on face-value heavily invested in generating data for global reporting, most of the country-level work funded can or is already strengthening national assessments for use, as in Ghana in the country cases. In seven of these countries support also includes working on better use of the data for policy, programming and instruction.

155. There is less evidence that separate global assessments are equally efficient investments. In four cases, D4FL supports separate global assessments (three ALMA AMPL assessments and one UNICEF MICS).⁴³ These are being done with national education authorities in all four, but the scope of work includes explicit capacity building objectives in only two, and we have evidence from two country cases that their impact on country assessment capacity is likely to be minimal (see Finding 3). Despite not contributing to the D4FL *Generate* objective, the results of these assessments could still be used at country level. While 12.7 percent of the UNICEF global budget is for advocacy on interventions, the strongest pathway to use of the three AMPLs is in Somalia, where the intent is to link AMPL data and results to the data and results of the Global Education Policy Dashboard by selecting an overlapping sample of schools across the two exercises. This would mean that data, for example on service delivery and policy, can be combined with learning data to enhance analysis and use.

156. While the D4FL support to the PAL Network to integrate the UIS criteria into its ICAN/ICAR assessments and conduct these in 13 countries,⁴⁴ also under the *Track* strand, does not strengthen country authorities' assessment capacity as such, it nonetheless strengthens in-country capacity in

⁴⁰ Seven criteria are proposed: (i) Alignment to the MPL and construct validity; (ii) Item content and quality; (iii) Population coverage and sampling; (iv) Assessment administration and data custodianship; (v) Reliability; (vi) Benchmark-based linking to the MPL; (vii) Maintaining standards over time

⁴¹ The GAML is the working group of the Education Data and Statistics Commission' (EDS Commission, formerly the Technical Cooperation Group on SDG 4 indicators) that focuses on learning data.

⁴² Djibouti, Ghana (2), Guyana, Kyrgyz Republic, LAO PDR, Malawi (2), Mozambique, Regional Caribbean, Rwanda, São Tomé and Príncipe, Sierra Leone. UNRWA is also supported to integrate the UIS criteria but does not intend to report on SDG 4.1.1. Mali is supported to integrate the GPF into its competency framework and for item bank development, but not to fully prepare an assessment.

⁴³ Nepal, Somalia, and Tanzania (ALMA), and Zimbabwe (UNICEF).

⁴⁴ Bangladesh, Botswana, India, Kenya, Mali (also ALMA), Mexico, Mozambique, Nepal (also ALMA), Nigeria, Pakistan, Senegal, Tanzania (also ALMA), Uganda.

civil society and academia which can be utilised by Government, and governments elsewhere. In addition, we know from the available financial data that 6 percent of the PAL Network budget from D4FL will be used for advocacy. This indicates some urgency to seek ways in which the existing work under *Track* and *Generate* can be leveraged better to support the *use* strand.

Finding 26. D4FL has invested in some relatively low-cost tools and solutions which are potentially scalable and replicable, and some that are more costly but more effective to build country assessment capacities.

157. Scalable and replicable approaches and tools can be seen as an indicator of efficiency because they have the potential to be used in other projects or contexts and encourage sustainability. While not all tracks or activities implemented by partners lend themselves to scalability and replicability in the same way, this MTE has unveiled some examples which are particularly relevant to this discussion.

158. D4FL is supporting the work of UNICEF to validate and finalise the FLM 2.0 module. The new module is intended to strengthen the measurement of learning outcomes within its MICS. The grant proposal lists 16 African countries where the module is likely to be rolled out as part of the new MICS exercise, in addition to the pilot in Zimbabwe which is directly supported by D4FL grant. Although its development was not supported by D4FL, AMPLs are similar in that they require less support at country level and could be replicated relatively easily.

159. Similarly, PAL Network's ICAN/ICAR assessments are likely to be perceived as more credible tools following D4FL funding and could attract additional funding to conduct similar exercises in other countries. The Network assessments are considered as cost-effective options for countries in need of data.⁴⁵ As discussed below, the PAL Network has already attracted funding from other donors to support data collection. In the future, this would be even more likely if data is accepted by UIS, something which PAL considers likely.

160. Other partners (ACER UK and the World Bank ALMA) have a strong focus on TA which is country-specific and therefore more difficult to replicate or scale. The TA capacity is scarce to deliver on this model for the remaining 51 LICs/LMICs with no or deficient data. In several of the ten LICs/LMICs and one region supported by the D4FL programme to report globally by 2026, difficulty recruiting appropriate TA has caused delays, especially in Francophone and Lusophone contexts.

161. In summary, while the global parallel assessments implemented at country level deliver nationally representative and globally comparable data and could be considered a relatively low-cost option for countries, it is less effective in building national assessment capacity than country-specific TA models. TA models, however, can be expensive, are difficult to scale and draw on relatively scarce expert capacity. The intention of ACER UK to make its tools available online for other actors to replicate, as well as to establish a community of practice for assessment authorities, could mitigate some of these barriers to the replication of this model.

162. The NFER model, if successful, represents a compelling third alternative, in which D4FL is supporting the establishment of the EARC, housed in the KNEC, with technical support from NFER and the British Council. This model seeks to make capacities that have been built in national

⁴⁵ See for example Results for Development, 2015.

assessment agencies available to other agencies in the region in a peer learning and exchange model. The focus of the EARC is currently less on integration of the UIS criteria for global reporting than on strengthening the assessment capacity of countries in general, though discussion about global reporting criteria has taken place, including at the EARC launch event.

Finding 27. The formative evaluation finding that crowding-in and leveraging other funding sources towards the achievement of D4FL outcomes remains valid.

163. The formative evaluation finding that the D4FL programme successfully crowds in (i.e. attracts or, at least, is in addition to resources from non-D4FL sources for the same activity) and leverages resources (i.e. the effectiveness and impact of D4FL-supported activities are supported by complementary activities not funded by D4FL) from non-D4FL funders remains valid. In combination, crowding-in and leveraging other resources contributes to maximizing the programme's reach and influence across the foundational learning landscape.

164. The **World Bank's ALMA and Accelerator programmes** are funded from the FLC Umbrella Trust Fund which combines resources from FCDO and other donors to support the implementation of the World Bank's foundational learning education programming. ALMA follows a catalytic approach, providing small, strategic grants to unlock larger investments in assessment system reform. Currently, ALMA grants are leveraging a total of US\$3.6 billion in World Bank active commitments to education projects in ALMA countries.⁴⁶ The Accelerator 2.0 programme utilizes its TA pool to leverage a more efficient and effective use of the significant resources that the World Bank provides to recipient countries (total education World Bank active education project commitments in Accelerator countries adds up to US\$4.1 billion, of which US\$3.05 billion are foundational learning projects).⁴⁷ Between the two programmes D4FL funding leverages US\$5.4 billion, netting out for countries supported by both programmes. Similarly, though not directly co-funded, **ACER UK's** interventions support and leverage other investments in learning assessments made by the World Bank and various countries.

165. In the case of accountable grant partners, co-funding arrangements were formalized or utilized to expand project scope. UNICEF's project is co-funded by the Gates Foundation. The FCDO cost component totals GB£1,175,262, complemented by GB£783,999 contributed by the Gates Foundation. The Foundation also co-funds the Accelerator programme. UNICEF is also engaged in discussions with Norway and BMGF regarding funding for different phases of the Learning and Skills Development Lab. The **PAL Network** has successfully secured co-funding for its assessments from Echidna Giving and Hempel.

⁴⁶ The team calculated this amount by summing the education shares in lending and grants to countries receiving ALMA or Accelerator grants, for active projects on the World Bank project database ([link](#)). The team first filtered the database for all active projects for each country that are tagged for education, then downloaded the project data, and applied the shares for education as set out in the project data. The FLC Umbrella Trust Fund team reported that the ALMA programme leverages US\$8.7 billion in World Bank *and* government education funding.

⁴⁷ The total amount of active education projects (US\$4.1 billion) is as calculated by the team (see footnote 46). The US\$3.05 amount for foundational learning programming alone, is as reported by the FLC Umbrella Trust Fund team.

Finding 28. D4FL has been able to influence and monitor how partners are using resources, but this influence is weaker in the case of the World Bank. There is also room to improve consistency and utility of partners’ financial reporting.

166. D4FL has established mechanisms to monitor and influence the resource allocation and implementation strategies of its partners, though the degree of influence varies significantly depending on the funding mechanism.

167. For accountable grantees, FCDO maintains significant monitoring control and influence. These partners are required to submit regular reports, including quarterly narrative and financial updates which detail project challenges, deviations from planned activities, and proposed solutions. Reporting is complemented by meetings. The level of detail provided by some grantees in reports supports monitoring efforts. However, it also presents inconsistencies across partners. For example, ACER UK’s and the PAL Network’s reports provide detail, by country and/or type of input or activity, but organise the detail differently. Other partners do not provide the same level of detail.

168. D4FL’s influence and control are reported to be weaker concerning resources channelled through ALMA and Accelerator 2.0, which are programmes of the World Bank FLC Umbrella Trust Fund. The fund management arrangements follows World Bank Trust Fund Policies and are in accordance with the agreement between the FCDO and the World Bank for FCDO support. In effect, however, the D4FL Management Team has less control over the specific countries and activities implemented through these programmes. The trust fund model also follows different reporting requirements compared to accountable grants. For example, the financial reports encompass all Trust Fund funding and do not offer project-specific detail. This too is in accordance with World Bank Trust Fund policies and the D4FL/World Bank agreement. The reduced level of control is partially mitigated through narrative reports or updates submitted upon request, although the structure differs markedly from that applied to accountable grants.

169. Considering the discussion above on the difficulty of obtaining a detailed overview of spending across the different programme tracks, a more detailed and consistent approach to reporting, in particular in relation to D4FL accountable grantees, would allow FCDO to better track the use of funds and facilitate a more detailed analysis of spending trends across the three strands.

EQ10 To what extent are partners delivering their activities in an efficient way?

Finding 29. Delays have been reported by multiple partners, often due to factors beyond partners’ control, but no significant consequences for future implementation and results are expected.

170. Multiple D4FL partners have reported implementation delays. These delays can often be attributed to factors outside the control of partners. In the first two years of implementation these included instability arising from political changes, as well as bureaucratic hurdles like delays in securing ministry approvals or spending freezes.

Table 5 Typical delays in programme implementation

Partner	Delays
ALMA and Accelerator	Set-up delays in making the grants (as evidenced in the formative evaluation). At country level, delays in the procurement processes and government approvals to implement the grants, or delays due to administrative changes in government staff.

Partner	Delays
	For instance, Accelerator reported isolated delays in Ghana attributed to leadership changes at the National Council for Curriculum and Assessment (NaCCA) following the December 2024 elections. A few countries reported procurement delays, as suitably qualified firms were difficult to find. None of these delays were expected to have a significant impact on the delivery of grant objectives.
ACER UK	Delays largely due to external contextual factors. For instance, ACER UK’s scoping of work in Mozambique was put on hold due to political uncertainty following the national election in October 2024. Similarly, spending freezes in Ghana delayed operations while the situation was reassessed. Despite these challenges, TA deliverables were mostly reported on schedule.
UNICEF	Experienced delays due to internal organizational factors (budget freezes) and external factors in Zimbabwe. Impact was managed through the re-sequencing of field tests.
NFER	Delays due to slow pace of KNEC’s setup of the EARC and partner organisational restructuring.
PAL Network	Country contextual factors are the most significant cause of delays in setting up and data collection, such as political instability (e.g. in Nepal and Pakistan) and weather (heavy rains in Nepal). Bureaucratic and coordination issues also caused delays, e.g. aligning PAL’s work with national statistical guidance in Nepal and the national assessment in Kenya. In Tanzania, the issue was internal with key PAL members leaving the network.

Source: ALMA and Accelerator Progress Reports; ACER UK, UNICEF, NFER and PAL Network quarterly reports; partner KIIs and workshops.

Finding 30. The delays have affected budget execution, but partners mitigate the effects through adaptive management.

171. Partner budget execution rates vary from 10 to 77 percent (Table 6).

Table 6 Budget execution by partners mid-to-late 2025, GBP

Partner	Allocated (GBP)	Planned (GBP)	Budget execution ⁴⁸ (GBP)	Budget execution (%)	Contract time elapsed (%)	Date of report
WB - ALMA	10,000,000	2,202,742	228,587	10.4%	N/A	Jun-25
WB - Accelerator		2,748,738	1,015,192	36.9%	N/A	Jun-25
PAL Network	3,000,029	3,000,029	2,322,315	77.4%	55.4%	Sep-25
NFER	201,546	201,546	111,231	55.2%	52.7%	Sep-25
ACER UK	1,999,971	1,999,971	996,490	49.8%	59.5%	Sep-25
UNICEF	1,175,262	1,175,262	253,839	21.6%	13.1%	Jun-25

D4FL partner contracts and financial reports

172. The **ALMA** programme’s low rate is largely attributed to the delays in procurement processes which constitute the majority (around 80 percent) of the grant value. Where possible, in allocated grants, the World Bank has taken actions to mitigate the impact of delays. For example, in

⁴⁸ There are differences in how these figures should be interpreted. For the ALMA and Accelerator programmes, funding is provided through the FLC Umbrella Trust Fund, and this figure represents funds paid to grantees against existing contracts. For the other partners, the figures represent reported spending by the partners themselves.

the Kyrgyz Republic, ALMA shifted from individual to firm-based procurement to reduce delays. In Lao PDR, the World Bank re-scoped activities (Grade 3 only) due to budget limits while preserving quality.

173. Conversely, the **PAL Network** has achieved an execution rate of 77.4 percent (September 2025). This comparatively higher rate is explained by higher costs of assessment compared to the budget proposed four years ago. The risk of overspending and the potential impact on the country assessments has been mitigated through the leveraging of external funding sources (e.g. Echidna Giving), which has allowed the PAL Network to still cover thirteen countries.

174. Other partners show execution rates that are in line with the expected pace of implementation, nonetheless, they have also deployed active management processes to maintain efficiency address unforeseen circumstances and mitigate the potential impact of delays. **UNICEF** re-sequenced field tests following internal organizational budget freezes and external factors (Zimbabwe). Similarly, **NFER** adapted its roadmap sequencing due to partner organisational restructuring and an increase in scope requested by the KNEC. NFER also introduced regular tripartite meetings to manage uncertainty. **ACER UK** prioritized flexible virtual delivery and adjusted its country focus based on demand and capacity.

4.5 Impact

175. Impact refers to the extent to which achieved D4FL outcomes are likely to contribute to significant progress against SDG 4.1.1 and learning poverty reduction in developing countries.

Summary of findings on impact: While there is little evaluation evidence on progress toward addressing learning poverty at this stage, there are some positive indications about likely progress toward this impact. In addition, the Accelerator Programme is directly working on addressing foundational learning poverty reduction through interventions in curriculum, materials and instruction.

EQ11 What evidence is there that D4FL's results will lead to improved quality of education and reduced learning poverty?

Finding 31. The evaluation has little evidence on progress toward impact at this stage.

176. D4FL's investment is largely in tracking and generation of data, which will create impact if it creates pressure for action, or if data are used to design and monitor learning interventions. The data expected to be generated through the D4FL programme, however, are not yet available to assess whether data are used to apply pressure at global and country level for action on foundational learning. Moreover, as detailed in Finding 17 to Finding 19 there is only limited explicit investment in analysis, dissemination and use of data generated or improved with D4FL support and limited evidence of working beyond the technical partners to engage users of data.

Finding 32. However, there are some positive signals about (likely) progress toward impact.

177. The intention to use the data to apply pressure for action on foundational learning is there, e.g. by the PAL Network which has detailed stakeholder analysis and action plans on the use of future data in place, including a common social media package plan to reach more audiences in relevant countries.⁴⁹ Capacity building of assessment authorities for better data is built into the

⁴⁹ PAL Network Advocacy Strategy.

scope of work in many countries, which increases the potential for impact. And in ten countries supported by ACER UK, ALMA and the Accelerator, activities that link generation of data to use of data is part of the scope of work, which is also a positive signal for potential impact. How these different components combine to support likely impact in Ghana, is discussed in Box 5 below.

Box 5 In Ghana ownership, analysis, and use of data support likely impact

In Ghana, there is clear commitment to improve foundational learning: it is included in the Education Sector Plan (ESP), is the focus of the new partnership compact, and a commitment voiced by ministers. National assessment data from the 2022/3 P4 assessment (pre-D4FL) were used by the Ghana National Association of Teachers on capacity building programmes and for research, as well as in Ghana Education Outcomes Project (GEOP) and GALOP evaluations, tied to performance-based funding with accountability pressure on schools and districts to improve foundational learning. Some optimism was expressed that the availability of NST data will lead to pressure on actors for better outcomes, while doubts were also raised, in particular whether there will be willingness to use and publicize the data if they show poor performance. Another question is to which extent the data are in practice openly available for use or analysis.

Source: Ghana country case study interviews, ACER UK and ALMA D4FL reports, country documentation

178. It cannot be assumed for all countries that support is likely to translate into impact, For example, in Somalia which is a fragile context in a protracted crisis with an assessment system that is only nascent, it is possible to foresee many risks derailing impact (Box 6 below). This does not make the Somalia support less important, however, given the D4FL's interest in how to support better learning data in such contexts.

Box 6 In Somalia, a longer and more difficult road to impact is likely

In Somalia, improving foundational learning is included as a priority in the ESP, and the development of a National Learning Assessment Framework (NLAF) is intended to guide future donor coordination and investment. This framework could in future be used to hold stakeholders accountable for foundational learning outcomes.⁵⁰ Due to the nascent state of institutions it will not be easy to ensure pressure on stakeholders to improve foundational learning in practice. The monitoring framework, an integral objective in the assessment framework might support this use by monitoring whether interventions on assessment and learning are on track. Overall, however, there are many hurdles between ACER UK's support for the development of the assessment framework and the use of the data as intended.

Source: Somalia country case study interviews and documentation review

Finding 33. In addition, the Accelerator Programme is directly working on addressing foundational learning poverty through interventions in curriculum, materials and instruction.

179. The World Bank Accelerator Programme has provided support to 11 countries on curriculum and instruction. Out of these, 10 have made significant progress in the last year, with support on track to be completed by 2026 (only Ghana is lagging).⁵¹ In six of the countries learning and teaching materials are supported, in four curriculum reviews, in three teacher training and professional development and in two remedial education.⁵²

180. Interventions are identified by World Bank staff in countries to fill critical gaps in curriculum and instruction. The World Bank Accelerator team has indicated that the identified interventions reflect alignment with country priorities. In Ghana, the only country case with Accelerator

⁵⁰ ACER UK, 2025

⁵¹ World Bank, 2025a. The Senegal Accelerator grant, which supported a local-language EGRA, is not counted here. Instead, it has been evaluated above as part of D4FL's data generation activities.

⁵² Ibid.

interventions in curriculum and instruction, respondents indeed confirmed that the programme was supporting important gaps in country knowledge, policies and systems for better learning outcomes. Respondents from the Ministry of Education confirmed that the Accelerator programme is covering important areas, and counterpart agencies and units were engaged in the work.

181. The evaluation has little triangulated evidence of the effectiveness of Accelerator country interventions. In, Ghana work has just begun, so results could not be observed. Interviews with the Accelerator management team regarding progress across the country grants, however, did not indicate any significant challenges in achieving what the grants set out to do.

182. At global level, support provided by the Accelerator includes two consultants to support World Bank country teams with expertise on foundational learning to ensure countries foundational learning borrowing is used well. Also, a Community of Practice has been established which holds quarterly workshops for World Bank teams, e.g. on language of instruction or teachers.⁵³ Furthermore, global goods on effective learning interventions that countries can draw on are being produced. These are demand-driven tools for World Bank teams to use to support key activities in preparation and implementation of World Bank-financed foundational learning programmes. Examples include an online course on book supply chain and procurement; a curriculum review toolkit (piloted in Nigeria); a rapid implementation survey to examine fidelity of science-of-reading pedagogy literacy programme implementation; and various notes and guidance materials (e.g. on EdTech, school visits, and integrated numeracy and literacy programme design).

4.6 Sustainability

183. Sustainability refers to the extent to which the net benefits of the D4FL continue or are likely to continue. At this mid-term stage, we look at indicators of potential sustainability, and factors that are likely to support and pose risks to lasting change.

Summary of findings on sustainability: There is evidence of promise regarding the sustainability of the D4FL's results at country- and global-level, but while the evaluation has evidence of several factors likely to support lasting change, it has also identified notable risks. Foundations for sustainability are being laid through strong government ownership, the institutionalisation of national assessments, and capacity-building efforts that are strengthening local systems and organisations. However, risks to sustainability include continued reliance on external funding and TA, particularly where item and test development and psychometric expertise and costly software are required; turnover of key personnel, particularly where TA focuses on individuals; and political economy dynamics – such as sensitivity to negative results or shifting priorities. Overall, while D4FL has set important building blocks for sustained benefits, long-term sustainability will depend on institutionalised resourcing, distributed capacity (beyond individuals) and continued incentives for countries to value and use foundational learning data beyond the D4FL window.

EQ12 Are in-country benefits likely to be sustained?

Finding 34. Overall, there are prospects for sustaining D4FL results, but these are uneven and highly fragile.

184. There are encouraging signs that many of the structural and policy changes supported through D4FL are on a path to sustainability, particularly where assessment systems were already emerging or where government ownership has been strong. Across multiple contexts, D4FL has

⁵³ World Bank, 2025b

supported national actors to take greater ownership of assessment agendas, and in some cases, to lead planning for future cycles with less external support. Indicators of potential sustainability are:

- **Government/local partner ownership and demand for data:** Somalia's NLAFF shows early signs of becoming a long-term policy instrument guiding coordinated assessment efforts across donors and federal states, with high government ownership of the framework, and commitment to develop a national assessment system.
- **Strengthening of institutionalised assessments:** In Ghana and Rwanda, national assessments (NSTs and LARS) are now embedded within education sector plans and wider reform programmes, with clear government leadership and institutional anchoring in NaCCA and NESA. In Ghana, the NaCCA was able to take a much larger leading role in designing the Primary (P) 2 assessment with integration of UIS criteria, following the support provided in 2024 on the P4 and P6 assessments.
- **Capacity building of local organisations (e.g. universities):** In Kenya, the establishment of the EARC and growing peer-to-peer learning mechanisms indicate progress towards institutionalising assessment capacity rather than one-off training. The PAL Network is providing centralised support to its member organisations who are implementing the ICAN/ICAR initiative, strengthening in-country capacity. Moreover, in countries like Nepal, implementing partners and their partner organisations described how the initiative is developing knowledge and skills around implementing assessment initiatives amongst the next generation by working with academic institutions.

185. However, the extent to which these changes will be sustained varies considerably by country and remains fragile in several contexts. Barriers to sustainability include:

- **Turnover of supported personnel:** In most countries, the strength of institutionalisation still relies heavily on a small number of motivated individuals, making progress vulnerable to staff turnover or political reshuffles. This was seen clearly in Somalia, where momentum slowed during the absence of the director of the national assessment unit. We also observed that implementing partners, in a number of cases, spoke of working with highly motivated and skilled individuals, and seeing positive results, but did not always note the risk of these individuals moving on and taking that capacity with them.
- **Funding:** Across countries, there is a reliance on donors and development partners to fund future data generation efforts. The evaluation has no evidence of supported government making commitments to fund future or data collection rounds from their own budgets. Even in high-ownership countries like Ghana, sustainability is not yet guaranteed: while structures and policies exist, recurrent funding lines are not secured.
- **Gaps in what is left behind:** In addition, several supported countries remain dependent on external technical expertise to meet the global reporting criteria and analyse learning data. Moreover, meeting UIS criteria can require specialist psychometric analysis and software. While free-to-access psychometric analysis software is available, in some settings paid licensed software has been used, reducing cost effectiveness and sustainability. This and other gaps in the "leave-behind" package (documentation, item banks, analysis code, standard operating procedures) pose risks to the feasibility and affordability of supported countries replicating data generation activities funded by D4FL.

Finding 35. There is evidence that D4FL may have contributed to increased demand for foundational learning data in some supported countries, though this demand remains fragile.

186. There is some evidence that D4FL has contributed to increased demand for foundational learning data in several supported countries. In Ghana, demand is strong and well-embedded, driven by the Education Sector Plan and performance-linked financing that requires learning data. Ministries, NaCCA, and partners are already using NST data to inform teacher development, school grants, and programme monitoring, and demand is expected to continue at least to 2030. Rwanda is using LARS data to inform curriculum and system reforms, and is planning further early-grade assessments and disaggregation. In Malawi, demand is reflected in the establishment of a National Assessment Taskforce, and in Somalia, the creation of the NLAf itself is a sign that foundational learning data are becoming valued for planning and monitoring. In Kenya, government directorates (as well as non-state actors) are increasingly seeking data and evidence from KNEC, and NFER reports strong interest in the EARC model across Kenya and the region.

187. Attribution to D4FL, however, varies by country, and in many cases the demand for data preceded D4FL and/or appears to be driven more by policy, donor incentives or external accountability requirements than by the programme alone. For example, in Ghana, demand for foundational learning data predates D4FL and is linked to World Bank Payment for Results incentives and Education Sector Plan commitments. Likewise, in Kenya and Malawi, government interest was present prior to support, making it difficult to distinguish D4FL-driven demand from existing momentum. In Somalia, while there is clear pressure from donors and Government for data, it is too early to determine whether D4FL's support has or will increase this demand.

188. Similarly, there is also evidence that the PAL Network model *could* generate demand for foundational learning data among non-state actors, in ways that can help maintain momentum. In Mexico, where national assessments had collapsed, the work of the PAL Network implementing partner, MIA, is viewed to have triggered "bottom-up" demand from teachers, state education authorities and civil society to reuse data for remediation, advocacy, and community engagement. PAL Network stakeholders believe that regional working and the credibility they have gained from aligning to UIS criteria will increase demand. However, while this non-state demand for evidence – and action – is the intended pathway to impact across the PAL Network countries, the extent to which it holds true for the D4FL-funded work remains to be seen once results are made available.

189. Factors threatening the perceived causal chain through D4FL activities to increased and sustained demand for foundational learning data include:

- **Political economy factors:** Partners have been aware since the outset that the demand for foundational learning data can be political. Leadership changes and the contested role of civil society, and the fear of "negative" findings are examples of political factors that can discourage generation, recognition and use of data. There is also a risk that demand for data is weakened once results become available. Evidence from other assessment initiatives shows that governments can be keen for learning data until the data show "negative findings", at which point they disengage or reject the methodology. The risk of this is particularly high where assessments have been developed outside of national systems. This may be partly offset by D4FL efforts to champion governments prioritising learning globally.

- **Limited engagement with and focus on evidence users to date:** As noted in Finding 6, D4FL-funded activities have primarily included engagement with technical departments within government who are responsible for learning assessment and data generation, rather than with evidence users. This risks data use – which will drive the demand for data – not becoming institutionalised, even where data generation capabilities are strengthened.
- **The risk of generating data that are not seen as useful:** Linked to this, D4FL-funded activities are less likely to generate future demand for data if the data initially generated are not seen as useful. This risk is more likely where evidence use has not been embedded in the planning of data generation activities from the start. The trade-offs between SDG comparability and local usability (see Finding 3) also come in to play here – if data generated are globally comparable but not useful for national or local decision-making, this is unlikely to generate demand for more data.
- **Short programming cycle:** In addition, the short programme cycle limits the opportunity to generate demand. Evaluation participants have expressed optimism that even if data show very low learning levels, there is a potential for government to see and tell a positive story if the data can demonstrate improvements over time. However, many data generation efforts funded by D4FL will only result in one round of data, limiting this potential and maximising the risk that government reject data because they perceive a “negative” story.
- **Perceptions about the feasibility of global reporting requirements:** Finally, there is also a risk that if D4FL-funded activities lead to data that are not accepted for global SDG 4.1.1 reporting (see Finding 21 above), this discourages governments from trying to generate data, if the global reporting criteria are seen as unreachable.

EQ13 What long-term changes in the global assessment landscape are likely to result from D4FL?

Finding 36. D4FL has played a convening role in important discussions that are believed to have led to international SDG reporting requirements that are expected to be stable and that are clear to global stakeholders.

190. Through D4FL, FCDO has engaged directly with UIS and convened discussions among UIS, other implementing partners and other education stakeholders that have contributed to the finalisation of the updated SDG 4.1.1 reporting criteria, and greater understanding of these criteria among key stakeholders in the sector.

191. Evaluation participants noted that the SDG 4.1.1 criteria are now more clearly articulated, and are expected to remain broadly stable going forward, even though periodic reviews are anticipated to ensure continued relevance.

192. In addition, it is hoped that D4FL has funded successful approaches to meeting these criteria that will serve as examples of how countries can achieve reporting against SDG 4.1.1 (see Finding 22). This has contributed to a growing perception that the criteria are becoming more usable and grounded in country realities, rather than purely normative requirements.

Finding 37. However, the feasibility and achievability of the criteria for most LMICs remains uncertain, and will depend on the interpretation of the criteria during the vetting process and how the global aid architecture moves forward.

193. The vetting process for determining whether a country's data meet UIS requirements is still to be confirmed and, as discussed in Finding 12, D4FL implementing partners expressed a lack of clarity about how the criteria will be applied and decisions about whether data will be accepted for SDG 4.1.1 reporting will be made. The level of stringency applied in the vetting process will have implications for how feasible and achievable the requirements are for LMIC/MICs, with knock-on effects on the sustained relevance of D4FL models for global reporting.

194. Affordability and access to TA are also central concerns for long-term feasibility. Implementing the criteria often requires specialist expertise, software and technical support, raising questions about whether countries will be able to comply without ongoing donor funding. Stakeholders pointed to the risk that compliance becomes dependent on external TA, potentially creating a model that is technically sound but not financially sustainable. The extent to which the criteria drive or sustain demand will therefore depend not only on their clarity and stability, but also on whether countries perceive them as worth the cost – in political, financial, and technical terms.

195. Ultimately, until the criteria are tested at scale across a wider range of contexts, their achievability and utility remain unproven.

Finding 38. There is emerging, but still limited, evidence that D4FL is beginning to stimulate increased demand for foundational learning data beyond the countries directly supported through the programme.

196. Some early signals suggest that D4FL-supported approaches are attracting interest from other governments and regional bodies. For example, Ghana's integration of UIS criteria into a national assessment has already generated follow-up enquiries from other countries seeking to learn from its experience, the World Bank ALMA team noted demand for support on data from countries not directly supported by D4FL, and implementing partners in Kenya and Somalia noted interest from neighbouring countries through peer-learning avenues such as regional networks and the East African Community. PAL Network partners also identified that ICAN/ICAR has sparked cross-country learning and informal uptake discussions.

197. There is also evidence that D4FL has contributed to normalising the expectation that LICs and LMICs can produce credible foundational learning data through multiple pathways—beyond traditional large-scale assessments. The programme has supported dialogue through platforms such as the Foundational Learning Exchange (FLEX), UIS convenings, and regional communities of practice, which have begun to shift perceptions among donors, technical agencies, and assessment bodies. UNICEF's plans to expand the use of harmonised MICS data for research and advocacy also indicate that D4FL may help influence demand indirectly, by increasing the visibility and policy relevance of foundational learning metrics at a global scale.

198. However, evidence of increased demand outside D4FL countries remains limited at this stage. Most examples relate to light-touch interest, peer learning or requests for information, rather than concrete commitments by non-D4FL countries to begin generating or reporting foundational learning data. In some cases, the demand appears to be driven more by regional or donor

pressures than by internal motivation. Furthermore, without clarity on UIS vetting processes and affordability of data generation, many countries may wait to see whether early adopters succeed.

199. Therefore, while D4FL has begun to catalyse early interest and cross-country curiosity, and is laying the groundwork for increased global demand, evidence of material, system-level demand beyond supported countries is still nascent. Further ripple effects are more likely to emerge once pioneer countries demonstrate sustained success and cost-feasible pathways.

5. Conclusions and Recommendations

5.1 Conclusions

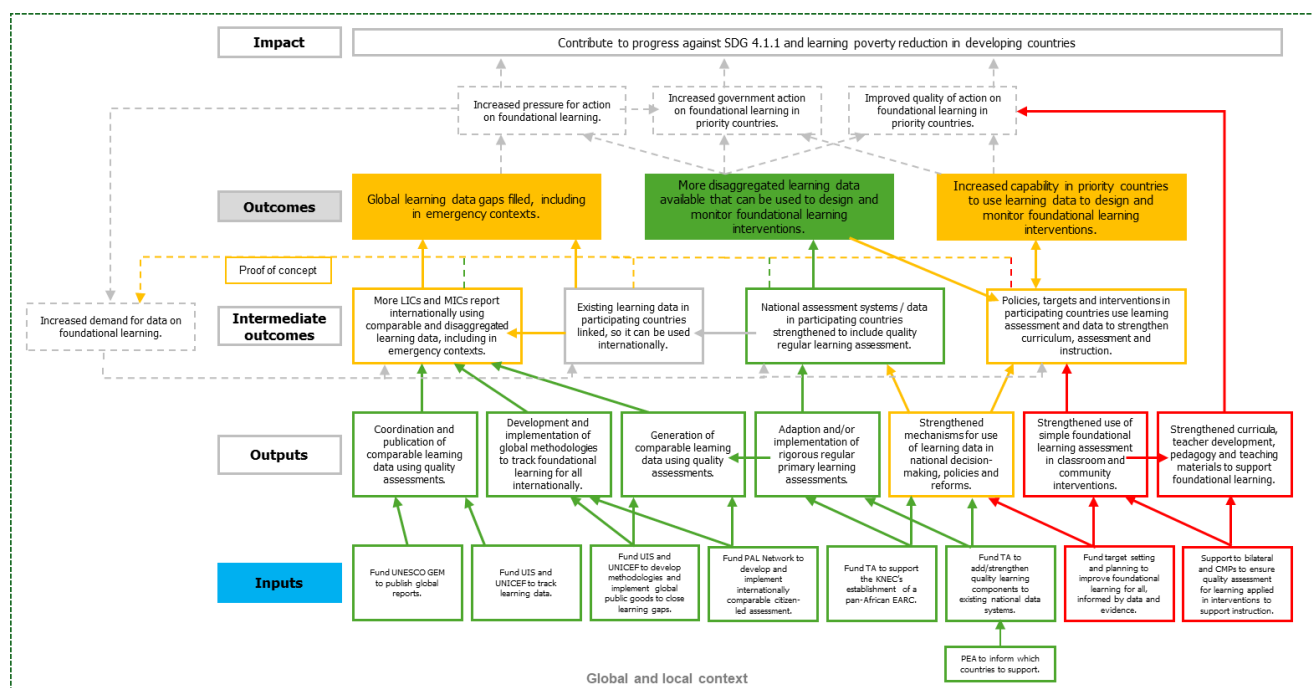
200. Here we present MTE conclusions against the two overarching evaluation questions: Is the D4FL programme a) achieving its intended results, and b) representing good value for money?

Is the D4FL programme producing the intended results?

201. Figure 4 below provides a RAG (red, amber, green) rating of the programmatic ToC, based on the findings of this MTE. Inputs, outputs, outcomes, assumptions and causal pathways are coloured green where they have been achieved, are very likely to be achieved, or very likely to hold true. They are coloured amber where they are at risk or likely to be only partially achieved or hold partially true, and are coloured red where they have not held true or are unlikely to be achieved. Grey indicates that is too early to draw conclusions on that element of the ToC.

202. The ToC shows that one intended outcome is likely to be achieved, whilst two are at risk or likely to be only partially achieved. As depicted in the ToC, and explained throughout this report, these anticipated results can be attributed to two factors: the investment in inputs to date, and the extent to which assumptions and theorised causal pathways have held.

Figure 4 RAG-rated ToC



- | Assumptions | Inputs to outputs | Outputs to outcomes | Outcomes to impact |
|--|--|--|---|
| 1. Sufficient funding is available. | 2. Learning Data Compact and other coordination mechanisms are effective. | 6. Global public goods developed provide solutions needed. | 15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundational learning. |
| 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. | 7. National governments are open to link assessment internationally. | 16. Governments and/or civil society have the resources (financial, human, technical) to (continue to) carry out learning assessment/data generation exercises. |
| 5. Targeted support to priority countries influences how data is used. | | 8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises. | 17. Governments respond to increased pressure for action on foundational learning. |
| | | 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. | |

Conclusion 1. Outcome 1, “Global learning data gaps filled, including in emergency contexts”, will likely be only partially met.

203. The programme will result in some global learning data gaps being filled, in that more internationally comparable data on foundational learning will be generated. It is yet to be seen, however, whether data generated through or as a result of support from the D4FL programme will be accepted by UIS for SDG reporting. Use of the data for SDG reporting is theorised to be important for increased pressure for action on foundational learning, leading to the intended impact of the programme: progress against SDG 4.1.1 and reduced learning poverty. While this causal pathway is potentially weakened by changes in the global education sector, there is still a risk that D4FL-supported data not being accepted further limits their potential for impact.

204. In addition, it is unlikely that the programme will make a notable difference to global learning data gaps for emergency contexts. If these gaps are not closed, this limits the potential for the programme to lead to impact on foundational learning outcomes for all children. This is particularly challenging at this point in time, because of the disruption to the aid architecture and progress being made on learning measurement in EiEPC contexts.

Conclusion 2. Outcome 2, “More disaggregated learning data available that can be used to design and monitor foundational learning interventions”, will likely be met.

205. The programme is expected to result in new, high-quality foundational learning data being generated, and national assessments in several countries being strengthened. These data will have

the potential to inform the design and monitoring of foundational learning interventions and policies. The potential for this to lead to more and better-quality action on foundational learning is strongest where data use has directly informed the design of assessments and assessment systems. But even where this is not the case, high-quality data on foundational learning levels at a general level can inform targeting and monitoring of interventions and policies at national, and in some instances, sub-national level.

206. TA strengthening government-led assessments will lead to improved data owned by government. Moreover, the evidence generated outside of government-led assessments is also expected to be available to government and other actors designing and monitoring foundational learning interventions. This includes public archiving of datasets (PAL Network) as well as publicly available disaggregated reporting (UNICEF).

Conclusion 3. Outcome 3, “Increased capability in priority countries to use learning data to design and monitor foundational learning interventions”, is at risk of not being met and, if it is met, will likely only be partially so.

207. The majority of D4FL-funded TA is focussed on assessment departments and technocrats, with the scope of support being primarily or exclusively focused on data generation and analysis rather than data use. These parts of government are not those responsible for designing, monitoring and improving programmes and policies. There are only two cases where D4FL-funded support is directed at strengthening the capacity of data users within government.

208. In addition, while data use activities are planned for much of the data generated or supported by D4FL, it is not clear whether and how those activities will strengthen the capabilities of participants to understand and make decisions based on learning data. If capability to use learning data is not improved in priority countries, the intended pathways to impact of increased government action and improved quality of action on foundational learning in priority countries are at risk. As a result, the programme’s potential for impact is reduced.

Is the D4FL programme representing good value for money?

209. The evaluation assesses good value for money as meaning that the D4FL’s objectives are relevant, are being achieved in a synergistic way, efficiently and sustainably, and with a high likelihood of achieving its intended impact.

Conclusion 4. In many respects, the D4FL programme presents good value for money.

210. Several evaluation findings point to D4FL presenting good value for money:

- In many of the countries it supports, D4FL is building assessment capacity while helping countries generate data to fill global learning data gaps. Where D4FL supports national authorities to generate data for SDG 4.1.1 through national assessments, it is also building assessment capacity. Its support for the PAL Network, which builds capacity outside of national authorities, and the NFER/KNEC/EARC support, is beginning to build capacities that authorities can draw on within and across country borders.
- In many cases D4FL partners respond to countries’ needs, and the MTE presents evidence that its results may be sustained. D4FL activities through ACER UK, the World Bank and NFER, for the most part, respond to the needs of supported country authorities. Where D4FL supports the generation of learning data through national assessments, country governments

have clear ownership of the data. In some cases, D4FL may also have contributed to increased demand for foundational learning data.

- D4FL has invested in piloting different methodologies to track foundational learning globally, including through integration in national assessments, through citizen-generated assessments, and through implementation of global assessment methodologies. The methodologies are replicable, especially AMPL, ICAN/ICAR and MICS. ACER UK and the PAL Network are intending to make their tools available for use by other actors, and the AMPL assessments used by the World Bank ALMA programme is available from UNESCO. This has the potential to leverage the D4FL investments into value for foundational learning across the globe.
- D4FL is providing relatively small amounts of catalytic funding that leverage much larger resources from elsewhere to generate more or better learning data, maximising the value of its investments. Partners have been efficient in the use of these resources, contributing further to value for money.
- D4FL has been complementary to other learning data initiatives, filling important gaps rather than duplicating efforts. Our evidence shows no duplication of existing learning efforts in countries. Rather D4FL-supported interventions fill important gaps in the learning data landscape, at times increasing the effectiveness of established mechanisms for learning outcomes.

Conclusion 5. Despite these strengths, several features of D4FL’s implementation reduce its value for money.

211. At the same time, there are factors that detract from the D4FL’s value for money:

- D4FL’s realisation of its *Use* objective is lagging. Achieving this objective in a synergistic way with its *Track* and *Generate* goals is critical for the programme realising its full value-for-money potential. Including a *Use* strand in the programme demonstrated the FCDO’s recognition that data generation cannot be isolated from data use, if it is to contribute to learning outcomes. In practice, only some D4FL learning data interventions include deliberate efforts to link generation to use, limiting the value of these investments.
- The D4FL programme also foresaw that complementary investments from other FCDO education programmes would help build the value of its investments in data generation. However, we have no evidence yet that this complementarity is realised in practice.
- Not all D4FL country-level investments succeed in contributing to the dual objectives of filling global learning data gaps and building national assessment capacities. In some cases, the D4FL investment will contribute to filling global learning data gaps, but with little evidence of impact on country capacity for assessment, demand for data or use of data. These cases are emblematic of the risk inherent in the programme design for better learning outcomes, when global reporting is prioritised over national use.
- In its current support to country assessment authorities for global reporting, D4FL relies on global TA capacity that may not be achievable or affordable for all countries. The support model that does build assessment capacity while supporting countries to report globally requires significant in-country support. The replicability of this model relies upon significant global-level capacity to support all remaining LICs and LMICs, as well as funding for such TA.
- Several contextual and programmatic factors weaken the sustainability of the programme’s results. Contextual factors include the lack of steady funding for assessments, turnover of supported personnel, and political economy factors that can discourage generation and use of

data. Shortfalls in D4FL implementation that weaken sustainability include gaps in what is left behind, limited engagement with evidence users, and the risk of generating data that are not seen as useful when evidence use is not embedded in the planning of data generation.

- Gaps in coherence and coordination between partners remain, risking competing approaches to data generation. All partners, other than the Accelerator initiative, share the objective of increasing the availability of quality data on foundational learning, but the level of focus on international reporting varies across partners and contexts and there is a lack of consensus about the importance of working within government systems. Moreover, in almost a quarter of the countries supported, D4FL is investing in data generation through multiple partners. While coordination has improved since the formative evaluation, gaps remain that risk incoherence and competing approaches to data generation.

5.2 Recommendations

212. The recommendation section aims to identify the changes FCDO and its implementing partners should make for the remainder of programme implementation to increase the likelihood of the D4FL programme delivering its results and realising value for money. We make five recommendations.

Recommendation 1. D4FL’s implementing partners should deepen the focus on data for in-country use in all cases of country-level support.

213. **Rationale:** D4FL has always contained potential trade-offs between generating globally comparable data for tracking purposes and generating more, and better, data for use in countries to improve learning outcomes. While both can lead to country-level action on foundational learning, the route to action from global reporting relies on the assumptions that countries will feel global pressure to improve learning outcomes when reporting comparable data, and that development partners will respond by funding interventions to improve outcomes. With changes in the global development context, where the sustainability of the global development and multilateral architecture is in doubt and where funding for global and country-level intervention is scarce, D4FL should be focused on ensuring that its support for data generation puts country use of the data at the centre. While many of D4FL’s data generation interventions aim to balance global reporting with national use objectives, there are cases where global priorities are put above national needs.

214. **Action:** In the data generation interventions that are still under way, implementing partners should work to ensure that country assessment capacity and national use of data are at the centre of the support, even when the initial reason for the support is reporting against SDG 4.1.1. This is done by ensuring ownership of the data through working through national authorities, rather than implementing parallel assessments; using technical assistance budgets to provide hands-on support and mentorship and address other national assessment needs, rather than focusing on the SDG4.1.1 data objective only; and building capacity development into the scope of work. This report touches on many cases where support was delivered in this way (e.g. Ghana, Malawi, Rwanda, Somalia and Rwanda to note a few). If any further country-level support is contracted, the scope of work should aim first to respond to country needs and build country assessment capacity, with the ability to report globally as the by-product, where it is on the agenda.

Recommendation 2. D4FL/FCDO should drive thinking about alternative options of supporting countries to build their assessment capacity.

215. **Rationale:** The MTE has concluded that the D4FL-funded model of international TA to build countries' national assessment capacity may not be fully replicable and sustainable in all LICs and LMICs requiring support, because of the cost and availability of such TA. Conversely, other models funded by D4FL, such as the implementation of AMPLs and MICS, require less targeted TA, but bring fewer benefits for country use of data. As a third model, D4FL's support to EARC has the potential to build a regional centre of expertise and peer learning for governments, although the relatively short delivery period for the support, as for all D4FL-funded activities, means the scope to demonstrate results within D4FL's funding period is limited, and the model still faces challenges to its success. FCDO and its D4FL programme partners are well placed to consider how support can be provided to help strengthen countries' data generation and data use capacities, while enabling global reporting, in a cost-effective and sustainable way.

216. **Action:** D4FL/FCDO and its partners should consider options for what cost-effective support to strengthen foundational learning data generation (and use) could look like in future. Taking includes thinking about *what* support might be provided, and *how*.

- Thinking about *how* support can be provided can draw on learning from D4FL implementation to date: for instance elements of the PAL Network approach (building local capacity outside of government), ACER UK's TA to the PAL Network (benefits amplified by benefitting multiple member organisations), and NFER's support to the EARC (regional peer learning and support). It might include thinking how FCDO and its partners could build capacity with a smaller number of institutions to undertake the highly technical components of learning assessment (under the mandate of Ministries of Education). Options could include building regional hubs, multi-government regional centres, private enterprises or non-government not-for-profit regional organisations.⁵⁴ Governments may also require support choosing from the approaches available to them, including such new approaches.
- Thinking about *what* can also draw on D4FL's experiences and other examples. It means considering which capacities are most crucial, cost-effective and sustainable to strengthen within government, and which capacities might be more efficient to outsource to the types of institutions described above.⁵⁵

217. FCDO can then consider what this means for investment: what support to fund for national and/or ultra-national institutions, and for state, multi-state or non-government organisations; which global actors to influence and how; and what investment to make in organisations that might provide that support.

Recommendation 3. D4FL partners should extend their efforts on ownership and use of data in their country-level support.

218. **Rationale:** This recommendation echoes the recommendation in the formative evaluation on working more politically to strengthen the use of data, and including teachers as a stakeholder group for D4FL *Generate* and *Use* activities.⁵⁶ The MTE has concluded that D4FL is still at risk of not

⁵⁴ For example, SACMEQ, CONFEMEN, Caribbean Examinations Council, the West African Examinations Council (WAEC).

⁵⁵ For example, governments in high income countries often outsource assessment development and benchmarking – e.g., examination boards in the UK and other nations, and ACER's implementation of national assessments in the UK and Australia (NAPLAN). In Rwanda, the National Examination and School Assessment Authority outsources administration of the national assessments to a firm.

⁵⁶ Recommendation 3 of the formative evaluation proposed to 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.

meeting, or only partially meeting, its *Use* objectives, which will constrain the extent to which its investments in data generation will lead to country action on foundational learning, ultimately reducing the impact of the programme. There is still time for D4FL and its partners to build stronger links to data use, where such links are not present in its investments in data generation.

219. **Action:** Four related actions are proposed:

- Over the remainder of the D4FL programme, implementing partners should see the extent to which they can work in countries beyond technical assessment units and with broader education authorities to instil more ownership of data to support its use.
- In any new support for data generation, country needs for data to inform policies, programming and instruction should lead, rather than lag, support for data generation.
- Any future support for data use should include more activities to support school managers and teachers to use assessment data.
- The D4FL Programme Team should work with fellow FCDO education CMP teams and country advisors to maximise the complementarity between the different strands of FCDO support, so that data that are generated with support from the D4FL programme, are used effectively.

Recommendation 4. The D4FL programme should seek ways to ensure that learning data are more inclusive, including more data for children affected by emergencies.

220. **Rationale:** This recommendation echoes a formative evaluation recommendation, on increasing the focus on inclusion of EiEPC settings as a primary objective of the programme.⁵⁷ Data for other often excluded groups, such as girls, OOSC, children affected by disabilities, rural children and others affected by drivers of exclusion are also targeted by the programme. While acknowledging that the D4FL programme partners have had limited ability to change whether often-excluded populations are included in the assessments the programme supports, and that exclusion is a global issue, the MTE has nonetheless concluded that the D4FL programme will deliver on these objectives to a limited extent only, precisely because there is little particular attention to the inclusiveness of supported assessments.

221. **Action:** The formative evaluation recommendation stands that the D4FL programme should increase its focus on the inclusion of children affected by emergencies. In addition, it should seek ways to help countries set up pathways to better availability of disaggregated data for children with disabilities and any other drivers of exclusion from assessment. This can be done in the following ways:

- The D4FL Programme Team and its partners should use their influence in assessment forums and on assessment actors to advocate for urgent continued action on global standards and open-source tools to assess children affected by EiEPC and disability. These remain areas where good practice approaches and frameworks at global level can provide an efficient solution benefiting all LICs/LMICs in future. This advocacy can draw on better understanding on barriers to inclusive data, in particular at country level but also global data, drawing on experience generated in the D4FL programme.

Recommendation 6 stated that partners should actively strengthen how they work politically to shift to regular assessment and use of data, and the FCDO and partners should invest in joint learning on how.

⁵⁷ Recommendation 4 of the formative evaluation proposed that the D4FL programme should increase its focus on the inclusion of EiEPC settings as a core programme objective.

- The D4FL Programme Team should arrange a workshop with the D4FL partners to discuss strategies for promoting more inclusive learning data on often-excluded children to countries, as part of its programme of regular meetings. The workshop — which can be facilitated by the evaluation team as part of its mandate for D4FL learning and work on strengthening the TOC accordingly towards the programme objectives — should include collating partners’ learning on country barriers to more inclusive data and their progress on approaches to generate more inclusive data outside of D4FL, for partners to provide information to countries on emerging global sources and tools.
- The D4FL partners should implement the agreed strategies over the remaining programming period, and report to D4FL Management Team on how they have done so.

Recommendation 5. The D4FL programme and its implementing partners should focus on the sustainability of results in the remaining programming period.

222. **Rationale:** The evaluation shows that the D4FL programme will achieve its objective of generating more disaggregated data that can be used to design and monitor foundational learning interventions, and may also contribute significantly to filling global learning data gaps. However, whether these results will be sustained beyond the end of the programme is at risk, from several contextual factors that the D4FL programme can still help mitigate (lack of steady funding for assessments, political change and staff turnover, and political economy factors influencing the demand for data), as well as programmatic factors that are more under partners’ control (such as what is left behind, limited engagement with evidence users, and the risk of generating data that is not seen as useful). There is still scope, in addition to what is already recommended in recommendations 1-4, to act to mitigate the risks and effects of some of these factors.

223. **Action:** In addition to the sustainability benefits from acting on recommendations 1 to 4, the D4FL programme and its partners should aim to strengthen the following aspects in their support for assessments:

- Budget institutionalisation of assessments (e.g., providing support for countries to consider co-financed transition plans; earmarked national budget lines for fieldwork/ analysis/ professional development)
- Building organisation-level capacity, rather than focusing support on motivated individuals (e.g., strengthening systems and standard operating procedures; provide advice on building assessment cadres/peer learning after end of TA)
- Ensuring that the leave-behind is robust (e.g., providing open or low-cost tools, codebooks, analysis scripts, item banks, manuals that provide institutional memory within countries, using open-source software, outlining options for replication across countries).

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

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

OFFICIAL

OFFICIAL

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

OFFICIAL

OFFICIAL

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](#)

OFFICIAL

OFFICIAL

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/>

OFFICIAL

OFFICIAL

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://www.unesco.org/en/evaluations)

OFFICIAL

OFFICIAL

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/dacriteriaforevaluatingdevelopmentassistance.htm>

OFFICIAL

OFFICIAL

- 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

OFFICIAL

OFFICIAL

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.

OFFICIAL

OFFICIAL

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.

OFFICIAL

OFFICIAL

- 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>

OFFICIAL

OFFICIAL

	<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

OFFICIAL

OFFICIAL

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.

OFFICIAL

OFFICIAL

- 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.

OFFICIAL

OFFICIAL

- 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.

OFFICIAL

OFFICIAL

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 \(unesco.org\)](#).

OFFICIAL

OFFICIAL

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

OFFICIAL

OFFICIAL

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).

OFFICIAL

OFFICIAL

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
Identity of the Controller and Processor for each Category of Data Subject	<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> <ol style="list-style-type: none">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:<ul style="list-style-type: none">• Names• Email addresses• Job titles

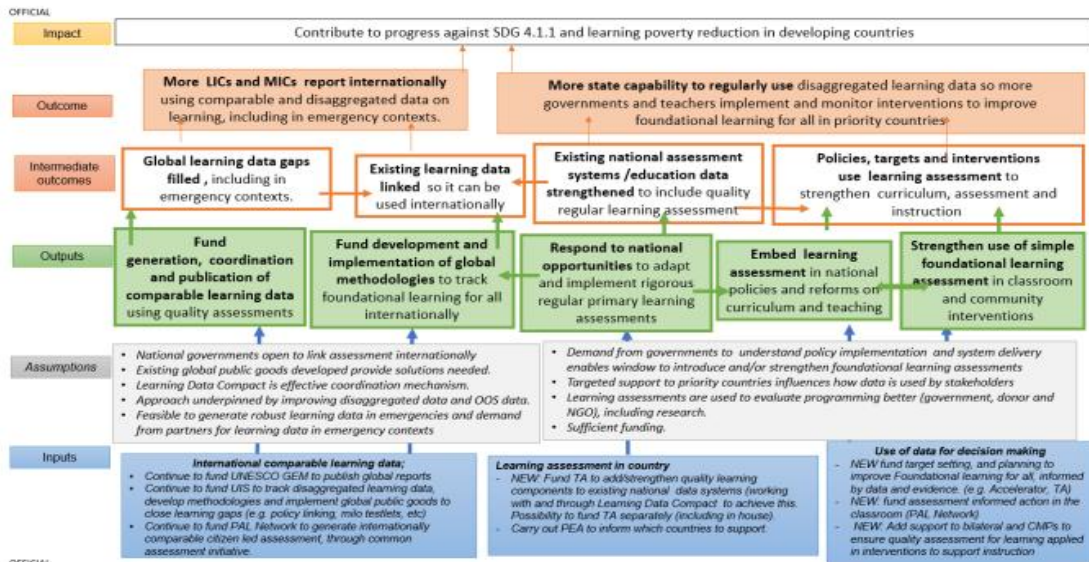
OFFICIAL

OFFICIAL

Annexes:

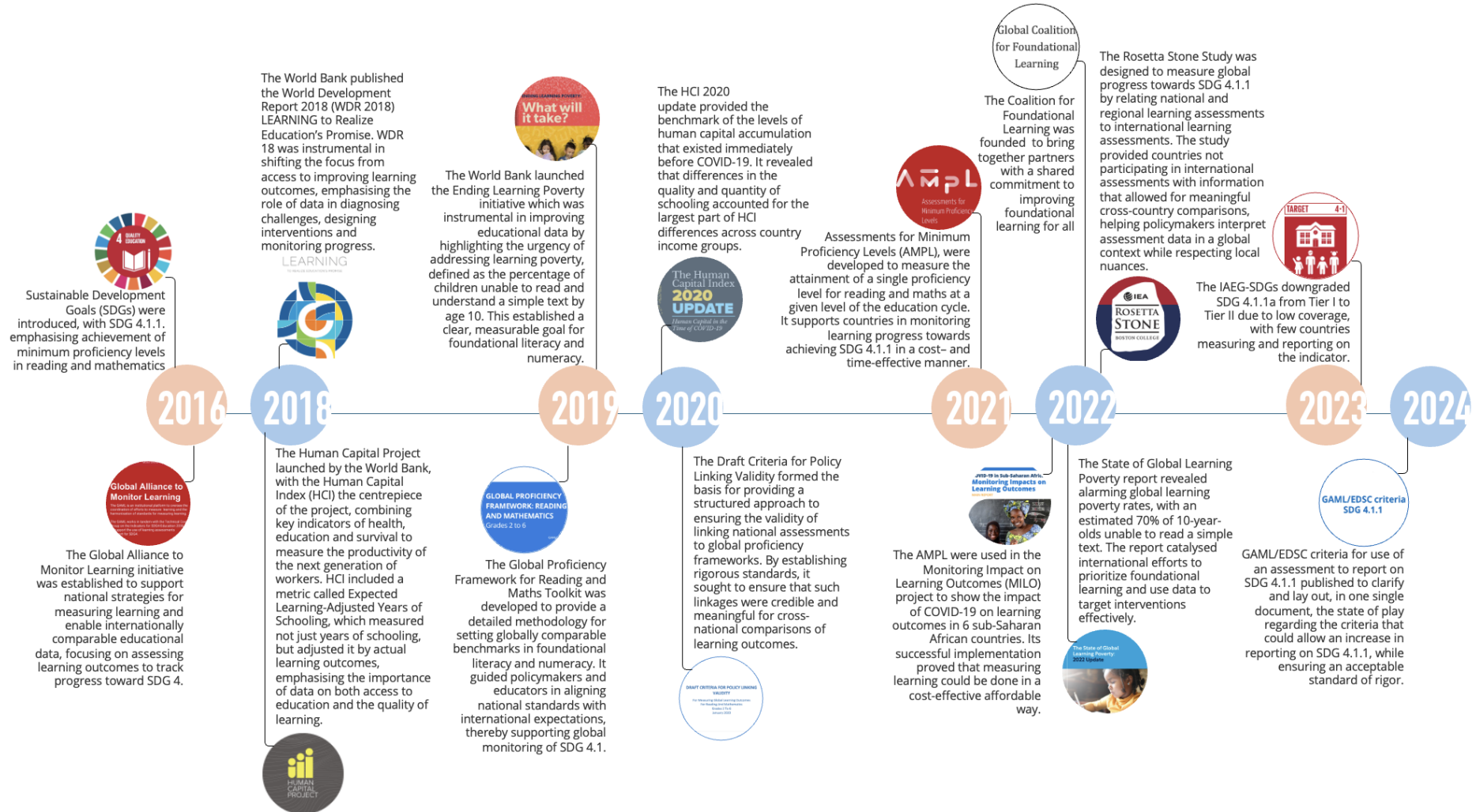
Annex 1: D4FL Logframe

Annex 2: D4FL Theory of Change



OFFICIAL

Annex 2 Timeline of Events in the Data for Foundational Learning Context



Annex 3 Education Assessment in the Context of SDG 4

Political and technical caveats of assessment of learning outcomes for SDG 4

1. SDG 4 aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.⁵⁸ 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.⁵⁹ This agreement was a significant milestone, given the varied priorities and challenges faced by different nations.
2. One of the most notable changes in SDG 4 in comparison to previous education goals is its shift in focus from access to quality in education.⁶⁰ Unlike former global education initiatives that primarily aimed to increase access, measuring school availability, enrolment, completion, etc., SDG 4 emphasizes the importance of ensuring that students are actually learning once they are in the classroom. By doing this, SDG 4 acknowledges a critical reality: merely attending school does not guarantee that children are actually learning.⁶¹ 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.
3. 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.
4. This political dimension was evident in the development of SDG 4. 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.⁶² As SDG 4 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).⁶³

⁵⁸ UNESCO, 2015

⁵⁹ UNESCO, 2017

⁶⁰ Sayed and Moriarty, 2020

⁶¹ Addey, 2024

⁶² Addey, 2024

⁶³ UNESCO, 2015

5. Specific targets, such as those mentioned above, come from the need for operationalizing SDG 4 into measurable outcomes, so that it can be monitored and assessed, ensuring accountability. SDG 4 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.⁶⁴ In the case of SDG 4, 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”.⁶⁵ 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.

6. 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 SDG 4 under a very vague indicator,⁶⁶ that makes tracking and accountability complicated, and thus the target easier to overlook.⁶⁷ 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.⁶⁸

7. 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.

⁶⁴ Sayed and Moriarty, 2020

⁶⁵ UNESCO, 2015

⁶⁶ 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.

⁶⁷ Sayed and Moriarty, 2020

⁶⁸ Addey, 2024

8. 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.

9. 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.

10. 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.

11. 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.⁶⁹

12. 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 and require significant technical capacity and expertise.⁷⁰ 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” able to inform educational policy has been questioned.⁷¹ 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.⁷²

13. 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

⁶⁹ Addey et al., 2017

⁷⁰ Wagner et al., 2018

⁷¹ Lockheed and Wagenmaker, 2013

⁷² Wagner et al., 2018

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.

14. Overall, we can see that while there is a formal consensus around SDG 4, 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 SDG 4 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 SDG 4 monitoring

15. 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 SDG 4 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.⁷³ 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

16. 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 2018.⁷⁴ This effort resulted in the global descriptor for MPL outlined in Table 1 of the main report.

Table 7 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	Students demonstrate skills in number sense,	ASER	6- to 14-y/o	Not applicable
			EGMA	Grade 1 to 3	Not applicable
			ERCE	Grade 3	Level 2

⁷³ UNESCO, 2015

⁷⁴ UNESCO, 2018

Domain	MPL		Equivalencies in cross-national assessments		
	2/3 Grade	End of primary	Assessment	Age/Grade	Level considered MPL
	and computation, reading simple data displays, shape recognition and spatial orientation.	computation, real world problems, basic measurement, 2D shape recognition, and reading and interpreting simple data displays.		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			

17. 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 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.⁷⁵

18. The UN SDG Extended Report 2025 (inputs as of 30 Apr 2025) reiterates that 4.1.1 reporting is drawn from international/regional LSA programs (e.g., PISA, PIRLS, TIMSS; PASEC, LLECE, SACMEQ, SEA-PLM, PILNA). It notes coverage remains stronger at end-primary and lower-secondary than in grades 2/3 and emphasizes UIS work on defining MPLs and linking.⁷⁶

19. 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 8. 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.⁷⁷

Table 8 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

20. In September 2024 UIS highlighted a major infusion of 235,596 national data points (54% SDG indicators); which bolstered SDG 4 learning and related series that feed 4.1.1 reporting.⁷⁸ In September 2025 UIS pushed a broad update to its education database (200+

⁷⁵ UNESCO, 2022

⁷⁶ UN, 2025

⁷⁷ Gustaffson, 2022

⁷⁸ UNESCO, 2024

countries) to support SDG 4 monitoring, with refreshed SDG 4.1 series accessible via the UIS Data Browser.⁷⁹

21. 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.⁸⁰ Data coming from national assessments often suffers from similar issues.

22. In response to data quality, reliability and alignment with MPL, the UIS has specified a set of 7 key reporting criteria, each containing a set of technical requirements that assessments must meet to be eligible to report against SDG 4.1.1. This document has been created to support the application of reporting criterion 1 – Alignment to the MPL and construct validity, by providing a clear definition of the MPLs and a blueprint for each educational stage – a) at the end of lower primary (Grades 2/3), b) at the end of primary, and c) at the end of lower secondary.⁸¹

Progress towards SDG 4.1.1 according to available data

23. Regarding progress, most countries are not on track to achieve SDG 4 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.⁸² 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.⁸³ However, these advances have been slowing down consistently, leading to a diagnosis of stagnation in progress towards indicator 4.1.1.

24. 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.⁸⁴ 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.⁸⁵ 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

⁷⁹ UNESCO, 2025

⁸⁰ Montoya, 2024

⁸¹ UIS and ACER UK, 2025

⁸² United Nations Economic and Social Council, 2023

⁸³ Gustaffson, 2022

⁸⁴ UNICEF, 2021

⁸⁵ Engzell et al., 2021

stress.⁸⁶ Four out of five of the 104 countries who studied the effects of the pandemic on their education systems experienced learning losses.⁸⁷ These effects are likely to be larger on countries that do not have the capacity to undertake studies such as these.

25. The 2025 Sustainable Development Goals report reported that most countries are off track to meet education targets for access, completion and learning outcomes. UIS emphasizes that many countries are now reporting, but the trend lines mostly show flat or declining proficiency once data are harmonized to the Global Proficiency Framework (GPF). Fewer than half of students worldwide reach the minimum proficiency level in reading or math at the end of primary (Indicator 4.1.1 b). Early-grade learning (4.1.1 a) remains the weakest area — foundational literacy and numeracy gains have stalled.⁸⁸

Pending challenges

Exclusion

26. 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.

27. The support to the PAL Network through the D4FL programme is a promising initiative to fill some of these gaps, as these assessments are household based.

28. In 2019 the Education in Emergencies (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.⁸⁹ However, in addition to the challenges faced across LICs and MICs, additional challenges exist in collecting data in EiE settings,⁹⁰ 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.

29. 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

⁸⁶ Minz and Chhabra, 2023

⁸⁷ United Nations Economic and Social Council, 2023

⁸⁸ UIS, 2025a

⁸⁹ Reductions in development funding may disrupt this architecture.

⁹⁰ Outhred et al., 2024

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.⁹¹

Uptake

30. 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.

31. 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.⁹² 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.⁹³ While ILSAs are methodologically sound, their global focus does not lend itself as easily to monitoring the immediate needs of local education systems.

32. 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.⁹⁴ 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.⁹⁵ 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

⁹¹ Johnstone et al., 2020

⁹² Tobin et al., 2016

⁹³ Howie, 2022; Trevino et al., 2016

⁹⁴ Kaffenberger, 2023

⁹⁵ Tobin et al., 2016

coherent integration of data, it is less likely that data on learning outcomes will result in policy changes.

33. Another issue affecting uptake is the quality of dissemination.⁹⁶ 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.⁹⁷ 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.⁹⁸ This misalignment between the intended message of the assessments and public perception can hinder the effective use of data in shaping education policy.

34. 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.⁹⁹

35. Despite these challenges, there is evidence of assessments influencing policy, albeit often through top-down changes. The areas most influenced by assessment results tend to be curricula and monitoring, evaluation and learning (MEL) 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.¹⁰⁰ However, changes in 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 and affordability

36. 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.¹⁰¹

37. Recent funding cuts in the sector will leave gaps in the budgets of LICs and LMICs to sustain their education programming, including gaps in funding for developing their assessment

⁹⁶ Creswell, 2017

⁹⁷ Tobin et al., 2016

⁹⁸ Grey and Morris, 2018

⁹⁹ Outhred et al., 2021

¹⁰⁰ Tobin et al., 2016

¹⁰¹ Görgens and Kusek, 2010

capacity and conducting assessments. The dismantling of US development support has been estimated to remove about US\$3-4 billion worth of funding from donor supported education programmes. Funding cuts by other major bilateral donors, including the FCDO, will create further funding pressures for governments.¹⁰² Ensuring assessment models are affordable and practical is especially important now.

38. 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.

¹⁰² Józwiak, G, 2025

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

1. In this Annex, Figure 5 sets out the original D4FL Programme Theory of Change (ToC). Please refer to Section 2.2 of this report for an explanation of the programme design and theory to accompany diagrammatic ToC.
2. Figure 6 (sets out an updated ToC for discussion. As explained in Section 3, we recommend adjusting 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.
3. We have proposed adjustments to the ToC to reflect all the currently intended D4FL programme inputs, outputs, intermediate outcomes, outcomes and impact, and how they are expected to lead to one another. Specifically, we have:
 - i. Set out three programme outcomes, reflecting the aims of the three programme strands (*Track, Generate* and *Use*)
 - ii. Rephrased the (now) third outcome so that capability to use data is not limited to state actors
 - iii. 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)
 - iv. 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
 - v. 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)
 - vi. Converted all outputs to programme outputs rather than FCDO actions (e.g. where outputs previously stated “fund...”)
 - vii. 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
 - viii. Added assumptions articulated by programme stakeholders in the formative evaluation, and indicated where in the ToC they apply
 - ix. We have also greyed out the one programme input that was planned but to date has not transpired.
 - x. In the following figure (Figure 7 to Figure 1) 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.
 - xi. At the start of the MTE in 2025, the D4FL programme partners were asked to review their ToCs, and a follow-up workshop was held with the World Bank FLC team. Workshops were also held with the UNICEF and NFER team to develop an explicated ToC for their contribution.

Figure 5 Original D4FL Programme Theory of Change

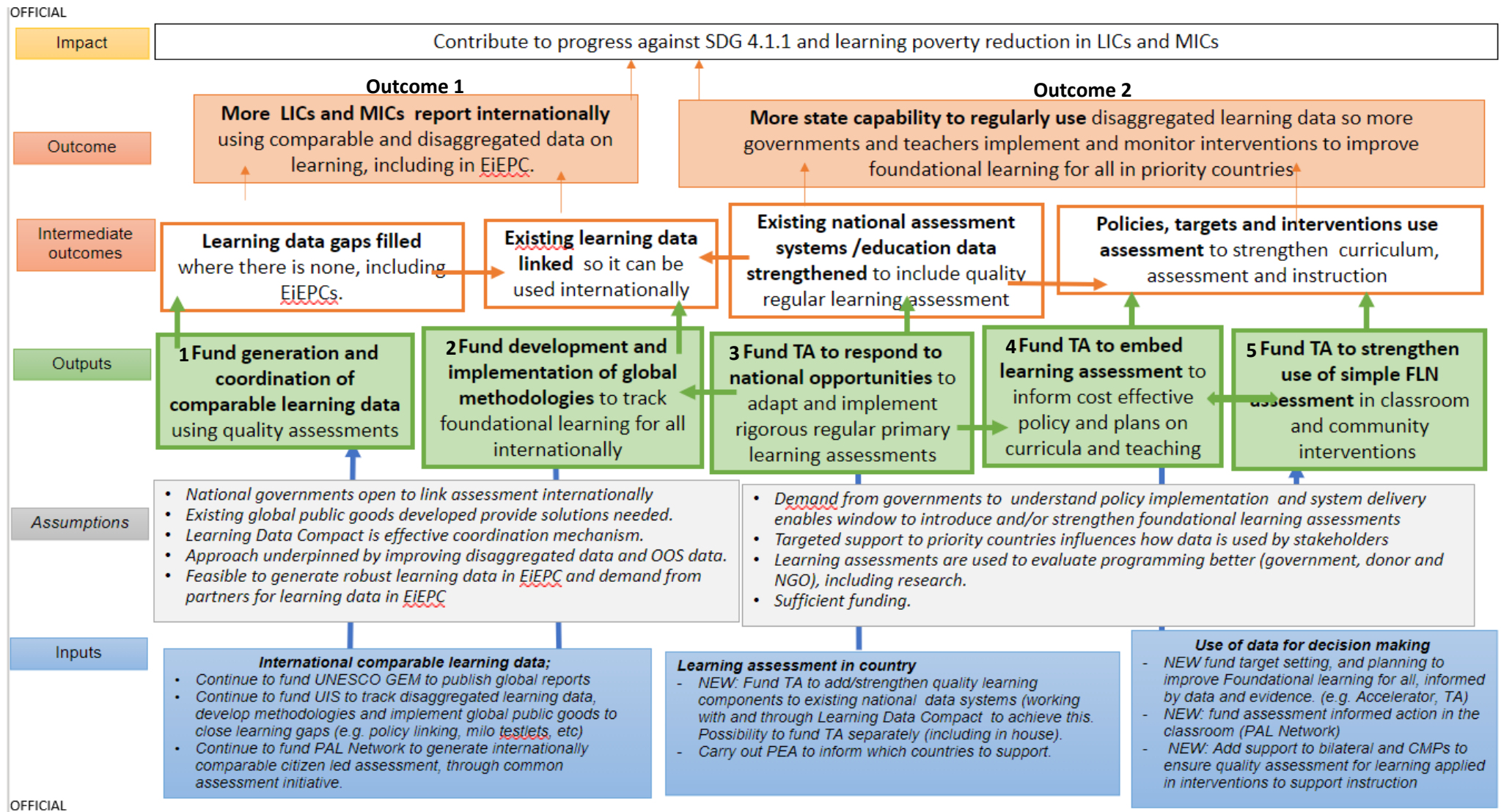
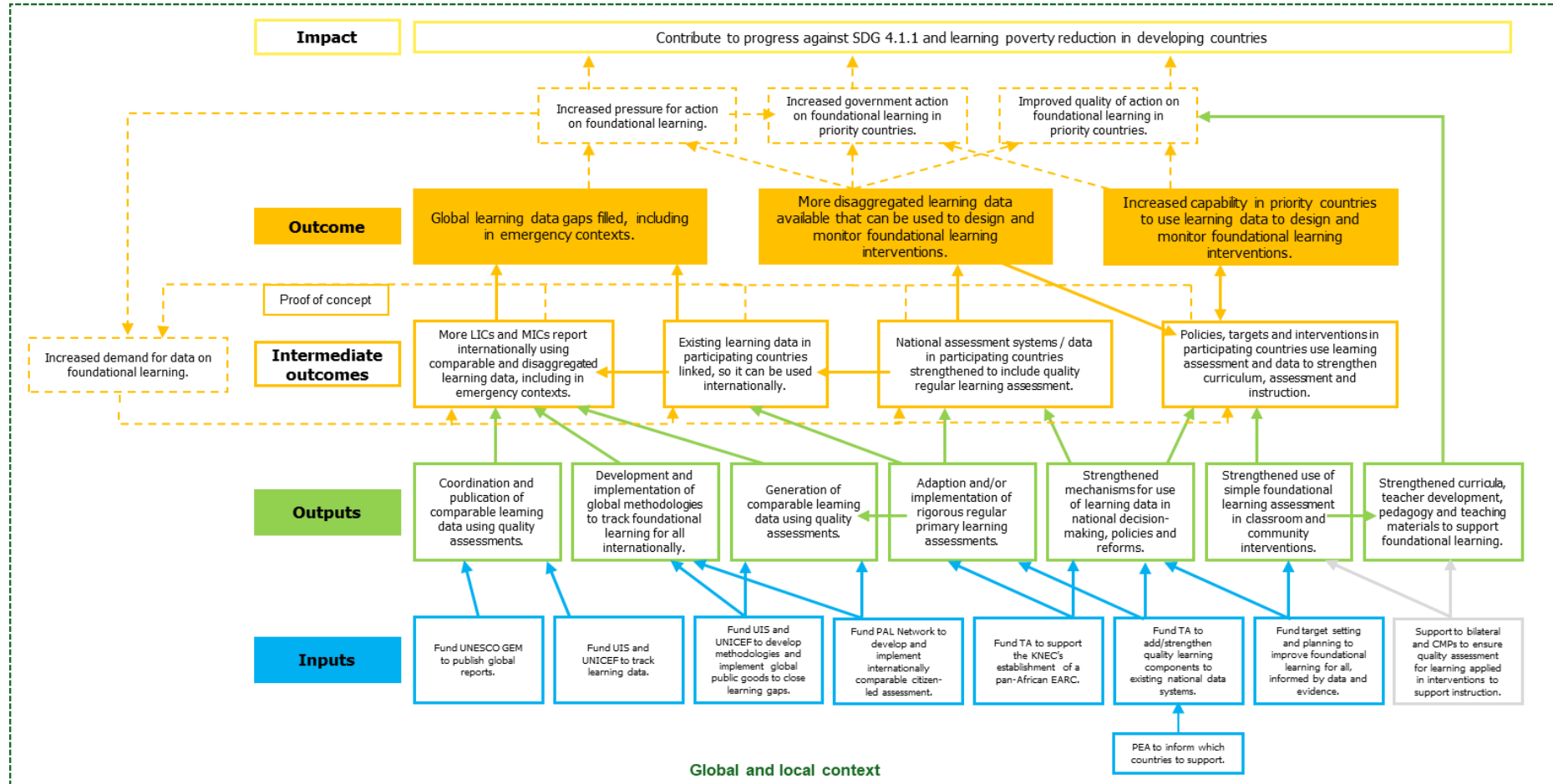


Figure 6 Updated D4FL Theory of Change



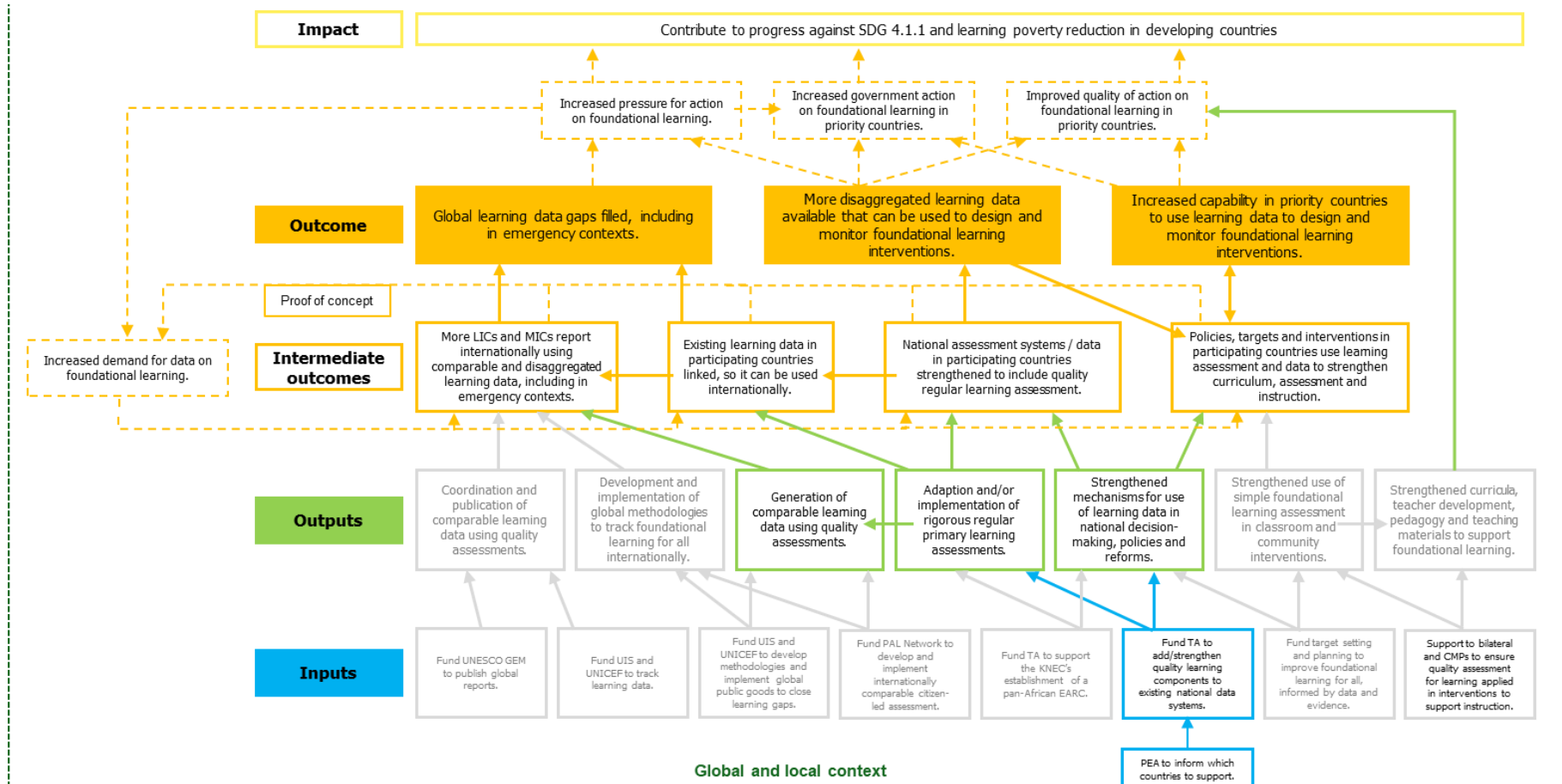
Assumptions

- Inputs to outputs**
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. Demand from governments to understand policy implementation and system delivery enables windows 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.

- Outcomes to impact**
15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundational 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 7 ACER UK ToC mapped onto the D4FL ToC



Assumptions

Inputs to outputs

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.

Outputs to outcomes

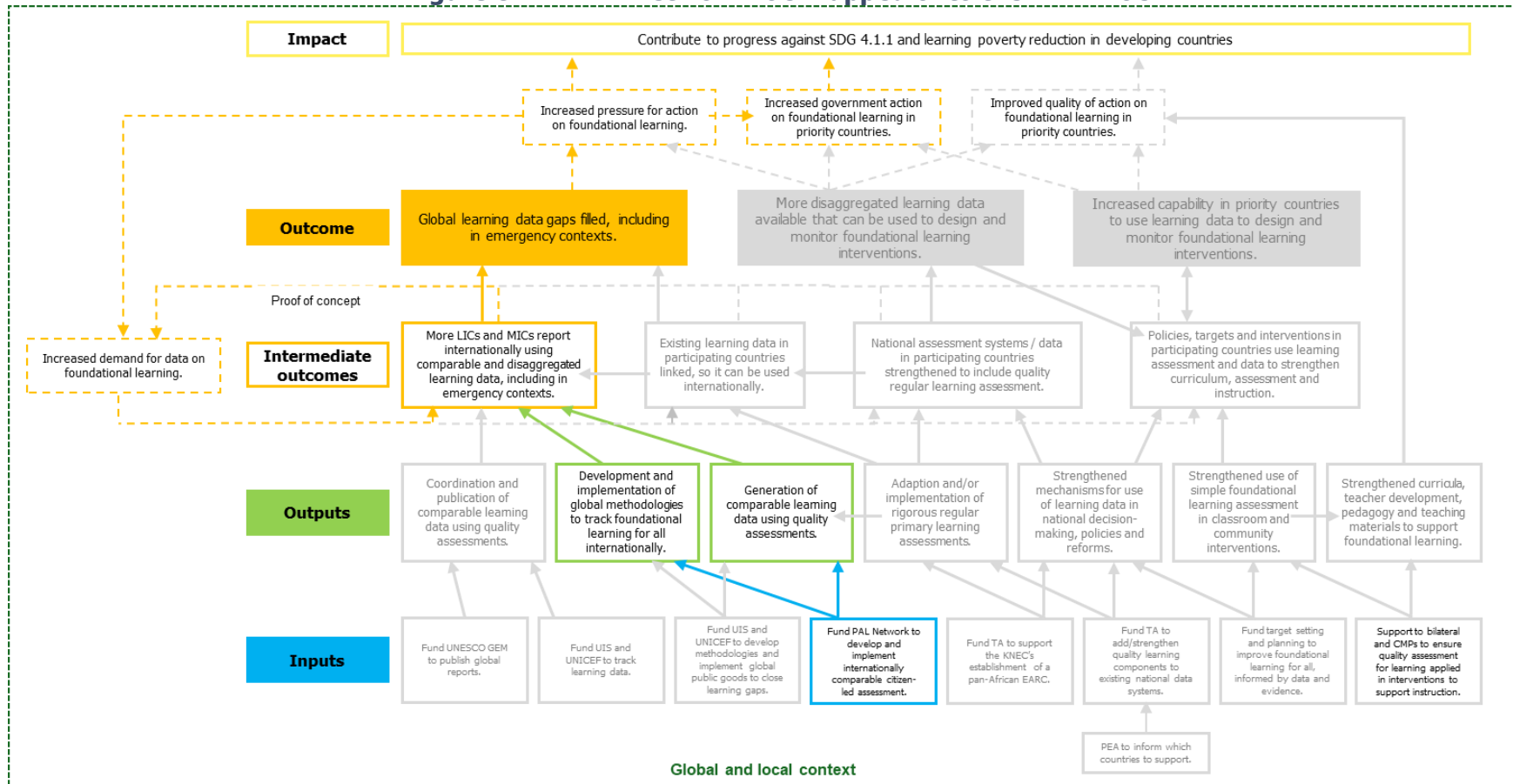
8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises.
9. Approach is underpinned by improving disaggregated data.
10. The approach is underpinned by improving out-of-school data.
11. It is feasible to generate 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 foundational 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 8 PAL Network ToC mapped onto the D4FL ToC



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.
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.

Inputs to outputs

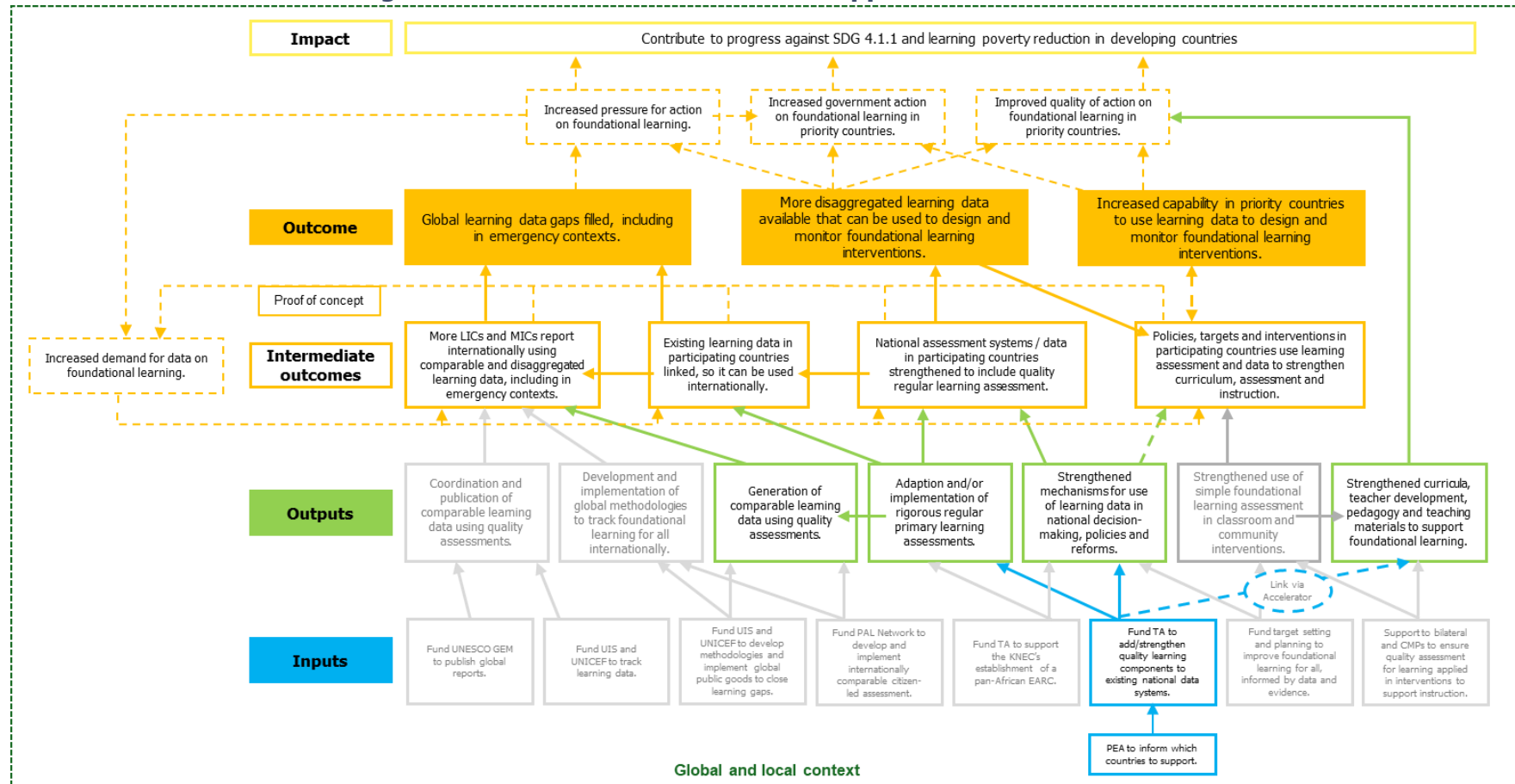
6. Global public goods developed provide solutions needed.
7. National governments are open to link/report assessment internationally.
8. Governments and/or civil society have the resources (financial, human, technical) to carry out learning assessment/data generation exercises.
9. Approach is underpinned 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.

Outputs to outcomes

15. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundational 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.

Outcomes to impact

Figure 9 World Bank ALMA ToC Mapped onto the D4FL ToC



Assumptions

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.

Inputs to outputs

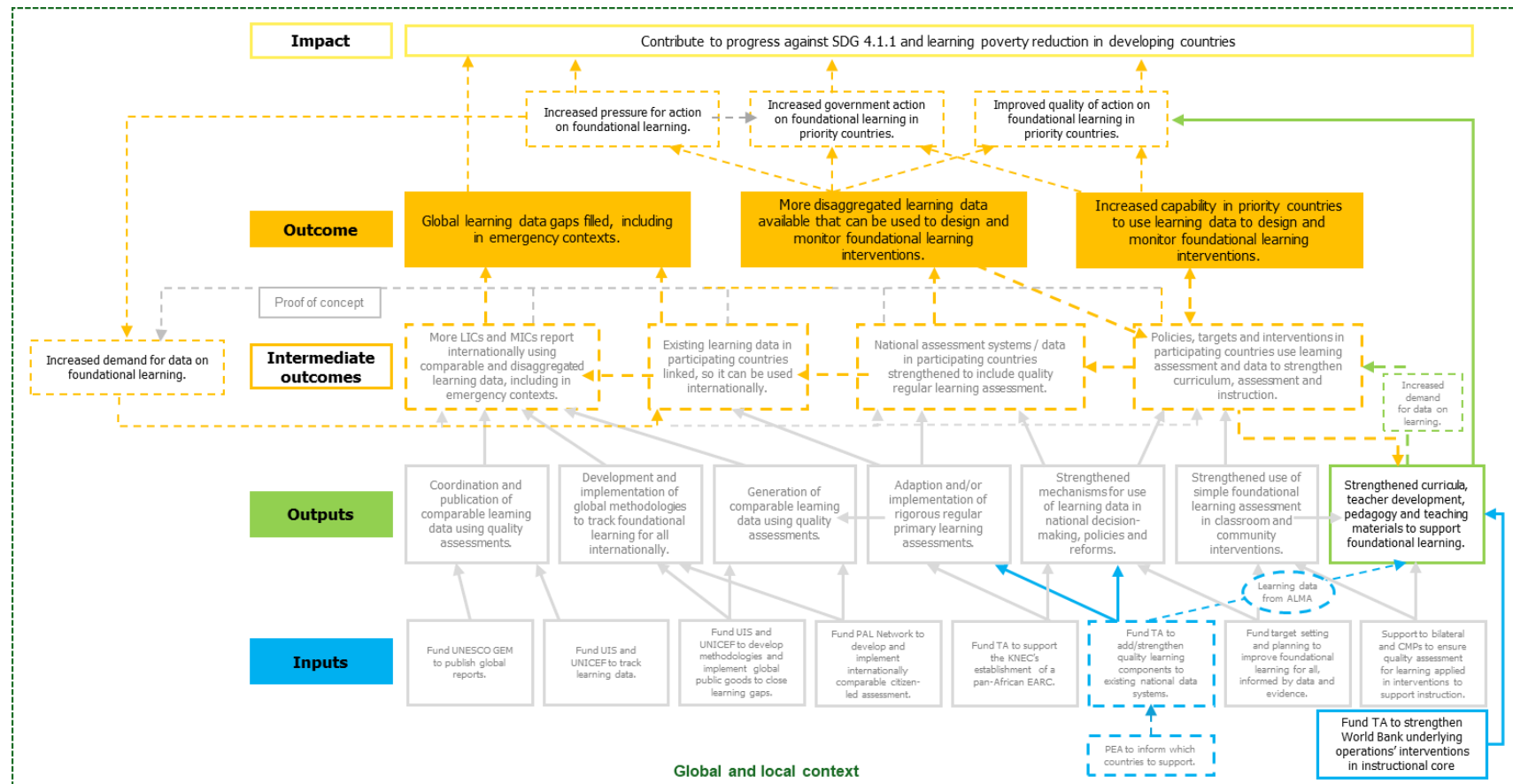
Outputs to outcomes

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 foundational 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 10 World Bank Accelerator ToC mapped onto the D4FL ToC



Assumptions

Inputs to outputs

Outputs to outcomes

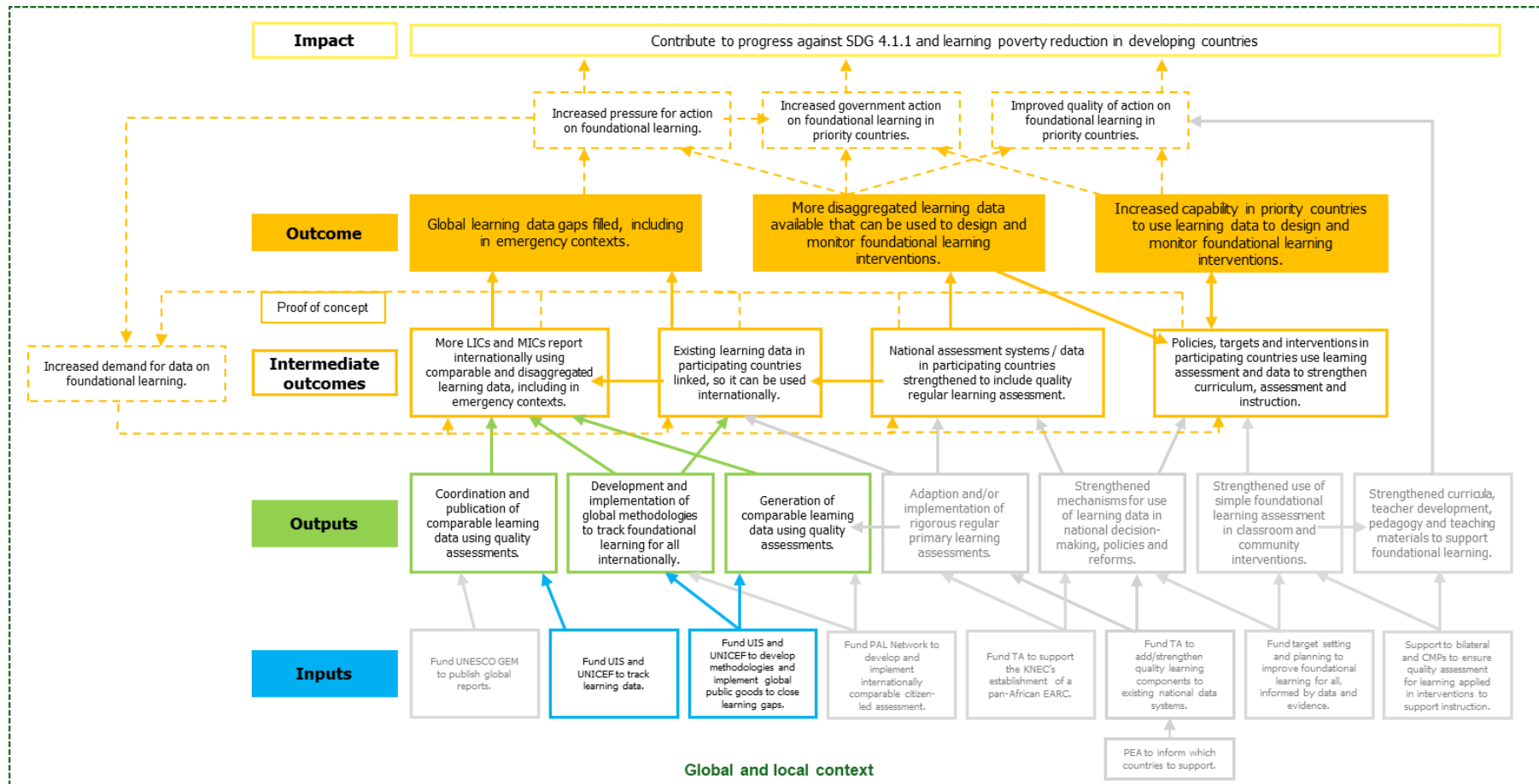
Outcomes to impact

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 foundational 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 11 UNICEF’s ToC mapped onto D4FL ToC



Assumptions

Inputs to outputs

Outputs to outcomes

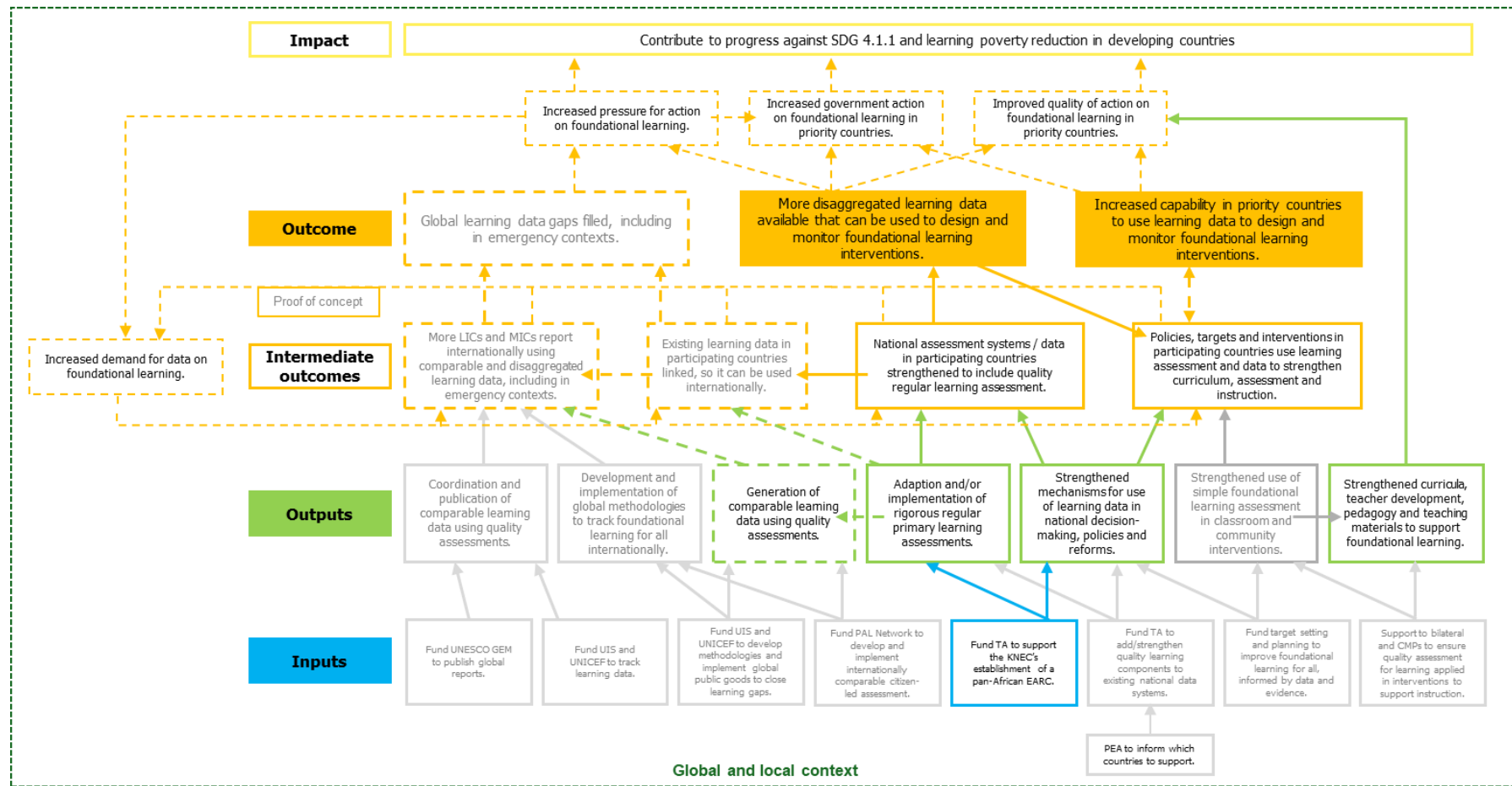
Outcomes to impact

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. Demand from governments to understand policy implementation and system delivery enables windows to introduce and/or strengthen foundational learning assessments.
5. Targeted support to priority countries influences how data is used.

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. Governments in priority countries have the resources (financial, human, technical) to improve provision for foundational 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 12 NFER ToC mapped onto the 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. 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 foundational 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 D4FL implementation overview

1. This annex provides a description of the D4FL Implementing partners' contracted work, and a summary of work done to date.

Table 9 D4FL Implementing Partners, Contracted Work and Work Done

Work contracted or agreed	Work done to date
ACER UK, contracted 20 November 2023, contract value £1,999,971	
<p>ACER UK is to:</p> <ul style="list-style-type: none"> • Support countries to use their national assessments for global reporting against SDG 4.1.1. • Strengthen learning assessment practices more generally in countries through capacity building • Improve use of learning assessment data within countries • Build a community of practice amongst LIC and LMIC on learning assessments <p>Work contracted in Ghana, Rwanda, Mozambique, Malawi, Somalia, Sierra Leone, PAL Network, and United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA)</p> <p>Work contracted to support the PAL Network on alignment with UIS criteria.</p>	<p>Ghana</p> <ul style="list-style-type: none"> • Completed final activities for NST 2024 reporting (SDG 4.1.1), including PCM benchmarking workshops (online and in-person) to set MPLs for reading and mathematics. • Prepared UIS reporting documents jointly with NaCCA • Delivered capacity-building training for NaCCA staff and SMEs on item writing, data analysis and standard maintenance; NaCCA now leading NST 2025 with limited ACER UK support. • Reviewed NST 2025 P2 English and Mathematics items against the Global Proficiency Framework; • Continued K2 assessment development; timeline extended to September 2025 for broader consultation; coordination ongoing with World Bank, FCDO, IPA and UNICEF. <p>Rwanda</p> <ul style="list-style-type: none"> • Supported NESA in finalising LARS 2025 field trial analysis, test construction and reporting to ensure SDG 4.1.1 eligibility • Conducted in-person workshops on test construction, data interpretation and contract management for large-scale assessments • Provided technical support on item analysis and selection for LARS 2025 test • Provided evidence and analysis for policy-making • Booklets and coding guides. • Following implementation of AMPL in June 2025, ACER UK coordinating with the AMPL team to align reporting timelines and mitigate data-linking risks. <p>Mozambique</p> <ul style="list-style-type: none"> • Scoping work halted due to political uncertainty after 2024 election: no work contracted. <p>Malawi</p> <ul style="list-style-type: none"> • Drafted and shared MoU with the Ministry of Education; awaiting approval. • Developed and circulated detailed project plan with timelines. • Began review of Priority 1 assessments (MLA, NRA, MLSS) with report due Aug 2025; Priority 2 review to start upon data access (CZT, DA). <p>Somalia</p> <ul style="list-style-type: none"> • QASD progressing with consultation on the national assessment framework; • Translation of outcomes into English pending; <p>Sierra Leone</p> <ul style="list-style-type: none"> • Project initiation completed <p>PAL Network</p> <ul style="list-style-type: none"> • Completed alignment review of ICAN and ICAR assessments (Criterion 1); • Planned PCM benchmarking exercise (Criterion 6); virtual meetings held and preparatory tasks underway. <p>UNRWA</p> <ul style="list-style-type: none"> • Worked commenced in July 2025

Work contracted or agreed	Work done to date
GEMR (not evaluated in full), contracted 03-Mar-23, contract value £3,200,000	
<p>GEM-R is contracted to provide high quality, policy-relevant research and analysis at global and regional levels draws on diverse methodologies and appropriately addresses core cross-cutting issues on SDG 4 progress; to produce publications and public goods that are accessibly communicated and synthesised for key audiences including policy makers; and to increase stakeholders’ engagement</p> <p>GEM Report contributed to the G7 Global Objectives report</p>	<ul style="list-style-type: none"> • Regional edition on school leadership and inclusion in Central and Eastern Europe, the Caucasus and Central Asia, Latin America • Second report in the SDGs series on nutrition and education • Gender report on education leadership and women • Third SDG 4 Scorecard report with a focus on the out-of-school rate (June) • Policy papers on aid to education (50) and aid to education in emergencies and protracted crises (51) • Regional edition on education leadership and digital transformation in East Asia drafted • Spotlight report research in five focus countries and background papers commissioned • Youth report on youth engagement in education legislation and policy making begun: two surveys administered to ministries of education and youth and student organizations • 2026 GEM Report on access and equity work advanced, including drafting 20 country cases, commissioning 3 systematic reviews, and preparing 159 PEER country profiles on financing for equity • SCOPE and VIEW website data updated on aid to education and out-of-school rates • New WIDE data on language and learning released and advocacy brief prepared • Maintaining the GEMR website • Four press releases and 52 events held
NFER, contracted Nov-24, contract value £200,000	
<p>NFER is contracted to undertake:</p> <p>Regional work to support the establishment and development of the Examination and Assessment Resource Centre (EARC) at the Kenya Examination Council (KNEC)</p> <p>Develop short, medium and long-term strategies for the EARC with</p> <ul style="list-style-type: none"> • Needs analysis conducted • EARC roadmap developed • Stakeholder engagement events conducted <p>Design and roll out a first phase of EARC training programme with</p> <ul style="list-style-type: none"> • Capacity transfer to KNEC/EARC (training of KNEC staff and trainers) • Core content for training programme developed • First round of trainings delivered 	<p>Launch of the EARC The EARC was officially launched on 4 July 2025 at KNEC, Nairobi, as a regional hub for assessment design, delivery, analysis, and use. Two demonstration sessions on foundational learning and competency-based assessment were held. Following the event, KNEC committed to expanding training into CPD programmes, formalising regional collaboration networks, developing an M&E framework, and securing sustainable funding.</p> <p>Capacity Building and Training Between August and September 2025, NFER delivered six of eight online train-the-trainer sessions for KNEC lead trainers, covering the full assessment cycle from framework design to reporting.</p> <p>Strategic Planning and Sustainability Work began on developing the EARC Roadmap under the British Council’s coordination. Terms of Reference were reviewed by KNEC and NFER, and consultant proposals were received by end-September 2025. Consultant selection and roadmap development are scheduled to commence in early Q4 2025 to guide EARC’s long-term direction and sustainability strategy.</p>

Work contracted or agreed	Work done to date
PAL Network, contracted 1 July 2023, contract value £2,800,000	
<p>PAL Network is contracted for the following:</p> <ul style="list-style-type: none"> • Common assessment model scaled to generate reliable data on children’s foundational learning that is comparable at the global level • Quality assured data collection of national level evidence using PAL common assessment tools on numeracy and literacy • Generation of a comprehensive global report with findings from national level data, which is directly comparable across countries at the global level • Engagement established with Multiple Stakeholders at Local, State, National, Regional and Global Levels <p>And to ensure alignment of ICAN and ICAR tools with International Standards, and build capacity for sustained implementation.</p>	<p>Outcome 1: Scaling the Common Assessment Model A comprehensive common assessment framework and item bank were developed and validated, with sampling and quality assurance protocols established to ensure nationally representative data. Country partnerships were formalized across 13 countries, and funding gaps fully covered through support from Echidna Giving. Field trials of the ICAN and ICAR assessments were conducted in all 13 countries, reaching over 7,000 children aged 5–16 across 18 languages. Standardized training, digital data collection systems, and monitoring protocols were implemented to ensure data integrity and comparability. Preparations are underway for in-person PMT training and large-scale national data collection.</p> <p>Outcome 2: Ensuring Alignment with International Standards The ICAN and ICAR tools were rigorously reviewed and revised with ACER UK to meet global reporting standards under SDG 4.1.1(a). Following iterative refinement, the UNESCO Institute for Statistics officially confirmed both tools’ alignment with the Global Proficiency Framework and Minimum Proficiency Levels, validating their use for global foundational learning measurement. Ongoing work with ACER UK includes preparation for a Pairwise Comparison Method workshop to strengthen benchmark linkage and ensure sustained compliance with international reporting standards.</p> <p>Outcome 3: Building Capacity for Sustained Implementation Extensive capacity-building structures have been established to support long-term national ownership of assessment processes. Country-level Project Management Teams have been trained through individualized sessions and virtual workshops covering sampling, data collection, and quality assurance. Guidance materials for translation, adaptation, and tool redevelopment were finalized, ensuring standardized implementation across diverse contexts. A strategic partnership with DataFirst (University of Cape Town) is enhancing data management, curation, and analytical skills while securing open access to PAL Network’s datasets.</p>
UIS (not evaluated in full), contracted 3 March 2023, contract value £4,400,000	
<p>The UIS is contracted to provide global leadership in the monitoring of SDG 4 Education Agenda enhanced through the development of internationally agreed norms and standards and the collection, production and distribution of timely and high-quality data</p> <p>Specifically, it should</p> <p>Establish the global SDG 4 data platform, with tools produced, regional mechanisms supported to monitor progress towards SDG 4 (with a focus on national benchmarks and learning data), and support provided to</p>	<ul style="list-style-type: none"> • Global Coalition for Foundational Learning: UIS-led efforts on learning progress monitoring and expanding data availability through new tools and methodologies. • AMPL-ab Project: Implemented in The Gambia, Kenya, Lesotho, and Zambia to measure lower and upper primary student proficiency. The report analysed SDG 4.1.1a and SDG 4.1.1b by gender across the four countries. Microsite with tools, guidelines, and publications developed, and AMPL visual identity created. • 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. • Releases: Education data released in March 2023 and September 2023.

Work contracted or agreed	Work done to date
<p>the G7 Global Objectives report.</p> <p>Support member states capabilities to design and implement the SDG 4 Education Agenda through strengthening capacity to monitor and report against SDG 4 targets</p> <p>Provide responses for relevant policy needs provided to develop the global education agenda based on research, foresight, innovation, and in collaboration with key partners</p>	<ul style="list-style-type: none"> SDG 4 Scorecard: Second edition launched in collaboration with the GEM Report during the Education Data and Statistics Conference.
<p>UNGEI / Cambridge Education (Not evaluated), Contract value £200,000</p>	
<p>Support to the G7 Global Objectives report</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.</p> <p>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>
<p>UNICEF, contracted 26 February 2025, contract value £1,175,262</p>	
<p>UNICEF is contracted for:</p> <p>Development of Validated MICS Modules to assess learning at lower and upper primary level (Output: Technical Validation and Harmonization)</p> <p>Increased Use of Learning Data in MICS for Advocacy (Output: Advocacy and Communication)</p>	<p>Pillar 1 — Validation and Finalisation of FLM 2.0</p> <ul style="list-style-type: none"> Completed second field test in Zimbabwe (Mutare and Mutasa districts), assessing 2,251 Grade 3–5 learners across 13 schools. Confirmed high reliability of reading and mathematics instruments; finalised FLM 2.0 forms (20 reading, 22 mathematics items). Developed adaptation/contextualisation templates and completed translations into Shona and Ndebele. Integrated FLM 2.0 into the MICS questionnaire, improving flow, language rules and stop rules. <p>Pillar 2 — Increased Use of MICS Data for Research and Advocacy</p> <ul style="list-style-type: none"> Established internal working group to expand advocacy based on MICS learning data. Initiated three key analyses: updated learning trajectories, age-for-grade outcomes, and learning inequalities. Completed preparatory work (methods, datasets, analytical plans); outputs scheduled for Oct–Dec 2025. <p>Pillar 3 — Harmonised Database and Learning & Skills Data Lab</p> <ul style="list-style-type: none"> Recruited consultant (onboarding Oct 2025) to lead data harmonisation across early-grade assessments. Began procurement for UX/UI design of the Learning & Skills Data Lab platform. Planned UX/UI workshop and system design activities (Oct–Nov 2025) to define platform architecture and governance.
<p>World Bank ALMA Programme, contracted 1 March 2024, contract value £5,000,000</p>	
<p>The ALMA Programme is contracted to strengthen national learning assessment systems in low- and middle-</p>	<p>Progress in country grants:</p> <p>Djibouti</p> <p>Djibouti has strengthened national assessment capacity through extensive technical support to the Ministry of Education’s Evaluation Department,</p>

Work contracted or agreed	Work done to date
<p>income countries to ensure that activities linked to the monitoring of student learning translate into meaningful improvements in education systems.</p> <p>The FCDO provides funding to the ALMA programme for small grants aimed at unlocking larger investments in assessment systems reforms. The grants are to be for activities ranging from developing national assessment frameworks, strengthening item banks, participating in regional and international large-scale assessment programs, and enhancing data use for policy-making, always with the overarching objective of reporting on SDG 4.1.1 a or b indicators</p>	<p>including workshops on report writing, sampling, psychometrics, and test anchoring. Staff also participated in regional PASEC training in Dakar. The country is implementing the 2025 PASEC cycle with partial ALMA support, aligning the Evaluation Nationale Intégrée with SDG 4.1.1 standards. Coordination continues with CONFEMEN and the World Bank to ensure methodological alignment.</p> <p>Ghana Ghana has advanced dissemination of its 2024 National Standardized Test results and begun preparations for the 2025 NST, including piloting plans for P2 assessments. Work has progressed on hiring a consultant and developing a national assessment portal to strengthen data use and reporting. Collaboration with ACER UK and the World Bank continues to align national assessments with UIS standards despite delays linked to post-election leadership changes.</p> <p>Guyana An independent consultant is conducting a comprehensive review of the National Grade Six Assessment (NGSA) to align it with SDG 4.1.1b criteria. The review includes system-wide analysis and recommendations for technical and institutional strengthening, with the final report expected mid-2025. Coordination with UIS and technical partners ensures alignment with international standards.</p> <p>Kyrgyz Republic The government has finalized the procurement of a technical firm to review and update the Grade 4 National Assessment for SDG 4.1.1 reporting. Coordination mechanisms with the Ministry of Education are in place, and a national expert roster has been established to support implementation. Activities will now move into the review and capacity-building phases.</p> <p>Lao PDR Following leadership changes at RIES, implementation has resumed with renewed government commitment. A consultant has begun technical work reviewing previous national assessments and developing new Grade 3 test items aligned with the curriculum framework. Collaboration with UNICEF and BEQUAL ensures alignment with broader education sector reforms.</p> <p>Malawi Preparations are underway for TA to support SACMEQ V data analysis and dissemination, with procurement of a consulting firm in progress. Coordination with ACER UK and FCDO continues to align activities and avoid duplication, supporting improved use of assessment data for policy and planning.</p> <p>Mali Data collection for PASEC 2024 has been largely completed, and work is advancing on aligning the national competency framework with international standards. Despite delays in some regions, the government expects to finalize data collection shortly. Next steps include validation of the framework and development of aligned assessment items.</p> <p>Nepal Nepal has approved the AMPL pilot, with translation and adaptation into Nepali completed by national consultants. A firm will be contracted to manage pilot data collection, informing the full-scale rollout in 2026. The World Bank and UNESCO continue to coordinate TA for implementation.</p> <p>Regional Caribbean The OECS has implemented quality assurance work with ACER UK to align early grade assessments in reading and math with SDG 4.1.1a standards. Revised tools are being administered, and momentum continues toward aligning Grade 4 assessments and benchmarking the Caribbean Primary Exit Assessment. Coordination under the WISTLE project ensures</p>

Work contracted or agreed	Work done to date
	<p>complementarity with regional partners.</p> <p>Sao Tome and Principe Preparations are complete for regional workshops on learning assessments and data use, with TORs developed for all supporting consultants. PASEC 2025 data collection is complete, and the next step involves hiring a firm to strengthen national assessments and improve reporting capacity. Coordination continues with CONFEMEN and lusophone partners.</p> <p>Somalia TORs for implementing AMPL have been developed, outlining administration, translation, sampling, and reporting processes. A firm will conduct data collection once schools reopen, integrated with the Global Education Policy Dashboard. Coordination with ACER UK, FCDO, and UIS is ongoing to manage implementation in a challenging security context.</p> <p>Tanzania Tanzania has finalized TORs for AMPL and shortlisted firms for procurement, with data collection scheduled for September 2025. UIS continues to provide technical guidance to ensure alignment with national and international reporting standards.</p>
<p>World Bank Accelerator Programme, contracted 1 March 2024, contract value £5,000,000</p>	
<p>The Accelerator 2 Programme (referred to elsewhere in the report just as Accelerator Programme) is contracted for:</p> <ul style="list-style-type: none"> Country grants to support the strengthening of instructional core and related delivery systems Foundational learning knowledge creation and exchange, including through producing global goods and facilitating demand-driven collaborations and knowledge-exchange 	<p>Central African Republic Progress focused on strengthening the national book supply chain and developing a textbook policy. Consultants completed a draft supply chain study and technical specifications for Grades 1–3 textbooks, and a Track and Trace system for textbook delivery was launched in September 2025.</p> <p>Côte d’Ivoire Implementation monitoring of the foundational learning reform has advanced through large-scale teacher surveys, classroom observations, and stakeholder consultations. Findings indicate improved learning outcomes but persistent material and training gaps. A report on teacher practices and recommendations is being finalized for dissemination.</p> <p>Ghana Work advanced on strengthening the literacy and numeracy curriculum and improving in-service teacher training. TORs were developed for international experts and a firm identified to evaluate the INSET model, though leadership transitions at NaCCA temporarily slowed progress.</p> <p>Eswatini The design and evaluation of a remedial education pilot were launched, with two firms contracted and preparations underway for implementation, with pilot delivery expected by June 2026.</p> <p>Liberia Curriculum revisions for early grade literacy were validated, and reviews of teaching materials are underway. Recruitment of specialists for numeracy and teacher professional development is in progress, with validation of literacy and numeracy revisions planned by October 2025.</p> <p>Madagascar A review of in-service teacher training and coaching systems has been completed, including surveys and interviews to inform redesign. The findings will guide the development of new teacher support tools following a nationwide survey scheduled for late 2025.</p> <p>Malawi The World Bank completed a numeracy curriculum review and began preparing for lesson observations to inform an updated teacher observation and CPD tool. Procurement of research partners is underway, with data collection planned for 2025 and outputs expected in early 2026.</p> <p>Republic of Congo</p>

Work contracted or agreed	Work done to date
	<p>Significant progress was made toward developing an “Essential Package for Instruction” through curriculum reforms, structured lesson planning, and capacity-building on literacy pedagogy. Plans now focus on adapting best practices to context, expanding training, and developing a comprehensive textbook strategy.</p> <p>Rwanda Remedial education materials were finalized, curriculum study findings from Kenya integrated, and a Track and Trace system deployed for textbook monitoring across 1,500 schools. Efforts continue to strengthen material use and finalize pilots of remedial education tools.</p> <p>Senegal The bilingual education program has expanded with new EGRA tools in six languages, completed baseline and midline assessments, and operationalized the Track and Trace system. Teacher training modules are being refined, and book tracking and further assessments are planned through late 2025.</p> <p>Sierra Leone The national structured literacy program scaled nationwide for Grades 1–2, with full sets of teaching and learning materials printed and distributed. Over 19,000 teachers and school leaders received training. A rapid implementation assessment is being designed to evaluate fidelity and learning outcomes, with data collection planned for early 2026.</p> <p>Tanzania Early grade teacher guides for literacy and numeracy were revised, shared, and are being digitized. Diagnostic tools have been contextualized for usability studies, with a 100-teacher classroom observation study planned to evaluate the updated guides’ effectiveness.</p>

Annex 6 Mid-Term Evaluation Methodology

This annex provides additional information on the methodology. It should be read in conjunction with Chapter 3 of the main report.

Evaluation Audience and Stakeholders

1. The FCDO is the recipient of the evaluation and the principal user. The D4FL Programme Team were also principal informants for the formative evaluation. The FCDO education advisors and broader GED team are also principal users of the evaluation, and in the case of education advisors, important contact points and informants of the evaluation specifically in the country cases.
2. Primary users and informants of the evaluation are the D4FL implementing partners, especially the World Bank, ACER UK, the PAL Network, NFER (contracted late in 2024) and UNICEF (contracted in March 2025). 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.
3. Other stakeholders in the global learning assessment architecture, such as the FCDO's peers in the Global Coalition for Foundational Learning, are also a primary audience and the evaluation conducted interviews with selected representatives from organisations in the Coalition not funded by the D4FL programme. The education authorities and other stakeholders, such as teachers and other assessment providers, in countries where the D4FL programme is active are also primary stakeholders, and were informants in case study countries.
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 TOR overarching questions for the mid-term round of the summative evaluation are (i) is the D4FL programme producing the intended results; and (ii) is it representing good value for money?
6. The MTE is answering these in conclusion, setting out the reasons why the programme did or did not produce the intended results and represent good value for money. We will evaluate value for money qualitatively, i.e. by assessing whether relevant objectives were achieved in a synergistic way, efficiently, and whether the results will be sustained and are likely to lead to the intended programme impact.
7. In Table 10 below we outline 12 high-level evaluation questions for the MTE, mapped onto the evaluation criteria and aligned to the explicated programme ToC. These questions are as proposed in the IR, except (i) for language changes to reflect the ongoing nature of the D4FL in the MTE or to make the questions clearer; (ii) rearrangement of questions against the coherence criterion; (iii) restructuring the effectiveness questions to reflect the change in the ToC.

Table 10 Summative evaluation questions (MTE)

Evaluation criteria	Evaluation questions as proposed in the IR
Relevance	1. To what extent does D4FL address the needs of LICs/LMICs?
	2. To what extent do D4FL's objectives and activities address the needs of the global community?
Coherence	3. To what extent is D4FL a coherent programme across strands and partners?
	4. To what extent are D4FL activities coherent and complementary with other global learning data efforts?
	5. To what extent is D4FL coherent and complementary with other FCDO programmes?
Effectiveness	6. To what extent is D4FL leading to the availability of more disaggregated learning data in LICs and LMICs that can be used to design and monitor foundational learning interventions?
	7. To what extent is D4FL leading to better capability in priority countries to use learning data to design and monitor foundational learning interventions?
	8. To what extent and how is D4FL filling global learning data gaps?
Efficiency	9. How well is D4FL allocating resources to maximise the chance that outcomes are achieved?
	10. To what extent are partners delivering their activities in an efficient way?
Impact	11. What evidence is there that D4FL's results will lead to improved quality of education and reduced learning poverty?
Sustainability	12. Are in-country benefits likely to be sustained?
	13. What long-term changes in the global assessment landscape are likely to result from D4FL?

Evaluation Theory of Change

14. The D4FL ToC anchors the summative evaluation, which will test the achievement of the ToC result pathways and validity of the assumptions. The evaluation will use an explicated version of the D4FL ToC (see Annex 4) an output of the formative evaluation. The explicated ToC drew on the theories of change discussed in the ToC workshops with partners along with broader learning about how the programme is intended to work.

Finalising the ToC

15. The formative evaluation report included several recommendations on refinement of the D4FL ToC for evaluation, including Recommendation 3 in the formative evaluation report, through which the team proposed that FCDO and its partners consider in a workshop in early 2025 how the D4FL can support strengthened instruction in classroom through its activities, as targeted in the original programme ToC. A follow-up ToC workshop, in the vein of the ToC workshops undertaken at the start of the formative evaluation, was organised and facilitated by the evaluation team, to tease out an approach and activities to support this area of the ToC in the D4FL programme.

Evaluation framework

16. The evaluation framework interprets the evaluation questions into sub-questions and assessment criteria to guide data collection and analysis. We provide the full framework for the MTE in Annex 7, outlining sub-questions, assessment criteria and data sources for each evaluation question. The language used in the framework reflects that the MTE is still occurring alongside D4FL programme implementation. We will adjust the framework in January 2026 as well to reflect the accountability orientation of the final summative evaluation, and to take into account lessons learnt about the questions, sub-questions and assessment criteria during the MTE.

Evaluation design and methodology

IR commitments mid evaluation design and methodology

17. The IR committed to selecting six countries for country cases, collecting data on the implementation by partners of D4FL-funded activities in context, in both the mid-term and the final summative evaluation. This commitment aimed to balance breadth of coverage of D4FL-funded activities, in different contexts, with collecting enough information in each context to draw findings on relevance, coherence, effectiveness of the activity.

18. The IR planned for one of the two data collections for each case study country being in-person fieldwork and the other desk work only. Countries were to be selected purposively to cover the variety of interventions different partners were implementing across countries, with the countries spanning continents, FCDO support scenarios and progress on learning assessments. The intent was that several cases would carry through from the formative evaluation. Data and information for D4FL countries that were not case countries, were to be analysed at the aggregate/portfolio-level only.

19. The IR planned an online survey of all stakeholders in all D4FL countries, in the mid-term and final evaluation. The survey was to follow country fieldwork, to test whether emerging themes apply in a wider number of countries. Countries not supported by D4FL were also to be surveyed, sharing some questions (e.g. on needs) with D4FL countries. These countries were included because of the relevance questions, to triangulate findings on the D4FL programme's relevance to LIC/LMIC needs from D4FL countries, avoiding positive selection bias in survey data.

20. Global-level data collection was to include interviews with global assessment stakeholders and desk-based document review.

21. The team committed to using similar approaches to analysis in the summative as in the formative evaluation. This included using the framework approach; rating D4FL performance as green, amber or red (RAG rating) against the evaluation framework assessment criteria; and using contribution analysis and a preliminary findings framework that assessed the strength of evidence to derive findings. We expected the summative evaluation to include secondary analysis of quantitative data sources, and descriptive analysis of the survey data to support findings.

22. Workshops with implementing partners were going to serve validation of emerging findings and help to explore any challenges, contradictions and implications for ongoing work.

23. As per the Terms of Reference, the IR proposed mid-term and final evaluation reports of no longer than 50 pages, excluding annexes, the production of an Evaluation Digest and an Evaluation Brief. The IR committed the team to active engagement on the Evaluation Reports, including the presentation of the final findings to the D4FL Management Team and other FCDO stakeholders, as

well as a webinar for secondary users. The IR also made commitments on how data will be treated in the evaluation, to ensure data protection and privacy.

Mid-term summative evaluation design

24. Table 11 below provides an overview of the design for the mid-term summative evaluation and indicates how the team used qualitative and quantitative data sources at country and global level to address the evaluation criteria and questions set out above.

Table 11 Evaluation matrix for the MTE

	Global level						Country level									
	Document review (D4FL, partners, global context)	External assessment expert interviews	Non-D4FL global assessment actors	Partner interviews and workshops	FCDO D4FL lead interviews and FCDO	Interviews with other FCDO centrally managed programmes	Document review	Data review	Government interviews	Partner interviews	Interviews with partners of other FCDO	External assessment stakeholder interviews	FCDO adviser interviews	Observations	Survey of D4FL countries	Survey of non-D4FL countries
Relevance																
1. To what extent does D4FL address the needs of LICs/LMICs?	X	X					X		X	X		X	X	X	X	X
2. To what extent do D4FL’s objectives and activities address the needs of the global community?	X	X	X	X	X											
Coherence																
3. To what extent is D4FL a coherent programme across strands and partners?	X			X	X				X	X		X	X			
4. To what extent are D4FL activities coherent and complementary with other global learning data efforts?	X	X	X	X	X		X		X	X		X	X		X	
5. To what extent is D4FL coherent and complementary with other FCDO programmes?	X			X	X	X				X	X		X			
Effectiveness																
6. To what extent is D4FL leading to the availability of more disaggregated learning data in LICs and MICs, that can be used to design and monitor foundational learning interventions?	X			X	X		X	X	X	X		X	X		X	
7. To what extent is D4FL leading to better capability in priority countries to use learning	X			X	X		X		X	X		X	X	X	X	

data to design and monitor foundational learning interventions?																
8. To what extent and how is D4FL filling global learning data gaps?	x	x	x	x	x	x										x
Efficiency																
9. How well is D4FL allocating resources to maximise the chance that outcomes were achieved?	x			x	x		x	x	x	x		x	x		x	
10. To what extent are partners delivering their activities in an efficient way?	x			x	x		x	x	x	x		x	x	x	x	
Impact																
11. What evidence is there that D4FL's results will lead to improved quality of education and reduced learning poverty?	x	x	x	x	x	x			x	x		x	x	x	x	x
Sustainability																
12. Are in-country benefits likely to be sustained?	x			x	x		x		x	x		x	x	x	x	
13. What long-term changes in the global assessment landscape are likely to result from D4FL?	x	x	x	x	x	x										x

Global-level data collection

25. At the global level, we researched the learning assessment and data context and the D4FL partners' D4FL-funded programmes, building on the document review and interviews we did during the inception phase. Specifically, we:

- Updated our document base for both the mid-term and final evaluation on countries reporting against SDG 4.1, primary documentation on assessment and learning data from global actors, and published literature on the global assessment landscape.
- Incorporated D4FL partners' submitted plans and reports, and sought to expand the financial data we have on D4FL programme spending at global and partner level, for the partners that are fully part of the evaluation, namely the World Bank, the PAL Network, ACER UK, ACER UK, UNICEF, and the NFER.
- Undertook interviews with global assessment stakeholders April to June 2025, especially with institutional actors, such as peers in the Global Coalition for Foundational Learning and the GAML, and funders and implementers of other assessment initiatives. We sought to expand who we talk to from the formative evaluation, but also had repeat interviews. We may do additional or repeat interviews with global assessment experts in 2026.
- Undertook interviews with the D4FL implementing partners' global teams. We interviewed the global teams at least once in the MTE, with follow-up interviews or workshops with the partners that are fully part of the evaluation towards the end of data collection.
- Met with the D4FL Programme Team during the evaluation for updates, and undertook a formal interview mid-way through data collection. We conducted an evaluation findings validation workshop with the D4FL Programme Team, followed by an extended meeting to discuss evaluation conclusions and recommendations.
- We conducted a workshop with the managers of key peer global education programmes (such as the WWHGE and SCALE, as well as Education Research in Conflict and Protracted Crisis (ERICC), Early Childhood Development at scale (THRIVE) and the Ed Tech Hub.
- We conducted a joined-up learning and reflection workshop with all the D4FL programme partners, and the D4FL Programme Team following the individual validation workshops.

26. Table 13 provides details on the KIIs and focus group discussions (FGDs) conducted.

Country cases

27. The IR said we would undertake six country cases in the summative evaluation, with a data collection for each country in each round. We also proposed taking as many as possible of the cases covered in the formative evaluation forward. Formative evaluation data collection, however, clarified that for some formative case countries an additional round of data collection in 2026 would be less valuable than adding additional countries, given the timeline for D4FL activities and the number of countries covered. Taking this into account, and the value of including additional countries, we proposed the following adjustment to the IR proposal in the MTE plan:

- We take all five countries covered in the formative evaluation forward in the MTE (Ghana, Kenya, Nepal, Senegal and Somalia).
- We keep three case countries from the formative evaluation in the sample for all three rounds of evaluation (most likely Kenya, Somalia, Senegal), but we swap out two (most likely Ghana and Nepal) for additional countries in the final summative evaluation.

- We leave the final decision on which country cases we will repeat and which we will replace, open until the January 2026 Summative Evaluation Plan update for the final round, depending on lessons learnt, progress made in countries where D4FL is implemented, and evaluation needs.
- We add a country for the mid-term and final summative evaluation to make up the sixth case.
- We undertake two in-person cases in 2025, and proposed four in 2026. This will allow in-person observations to be predominantly when the programme has advanced further. We will make a final decision on which cases will be done in-person in 2026, in January 2026, when we will have more information on all the cases.

28. Non-case study D4FL countries are analysed at the aggregate level, utilising information from global documentation and data, including the FCDO D4FL Annual Reviews, partner reports, the D4FL log-frame and publicly available documentation about countries that are reporting available data. They were also surveyed, as is explained below.

Country cases

29. Following learning from the formative evaluation and further conversations with FCDO, the team identified six countries for the MTE, namely Ghana, Kenya, Nepal, Senegal, and Somalia – the five countries included in the formative evaluation – as well as Mexico. Box 7 below explains the selection process in more detail.

Box 7 Country cases selection

The selection of country cases for the MTE of the D4FL programme followed a purposive sampling strategy, balancing typical and outlier cases. Six countries were selected: Ghana, Kenya, Mexico, Nepal, Senegal, and Somalia. These were chosen to reflect diverse contexts, partner involvement, and not least, continuation from the formative evaluation.

Key selection criteria included:

- **Representation of partners:** All partners must be covered by at least one case, but preferably two cases. For this purpose, ALMA and Accelerator are seen as two programmes. Two or more partners must be active in most countries. In at least one country only the PAL Network should be active (Mexico). We wanted to observe at least one country where ACER UK and the World Bank ALMA programme overlap (Ghana and Somalia), and one where ACER UK and Accelerator overlap (Malawi, proposed for 2026 inclusion). We proposed to replace Ghana with Malawi as a second case besides Ghana, where ALMA, ACER and Accelerator overlap. Senegal is also interesting because of the overlap between Accelerator and the PAL Network. We will include the NFER from the MTE in the Kenya case study, where its counterpart (KNEC) is based.
- **Overlap with other centrally managed FCDO programmes:** We wanted to include at least one country where the D4FL programme overlaps with country-level activities of the WWHGE and/or SCALE programmes. For WWHGE, D4FL overlaps in Kenya, Ghana and Pakistan, amongst the proposed case countries. For SCALE, Malawi (pilot already started), and Nepal and Somalia (possible pilots in the pipeline) overlap.
- **Overlap with UIS-supported AMPL countries:** D4FL supports the UIS to pilot and implement AMPL in countries. We wanted to include countries in the sample where the World Bank, ACER or the PAL Network supports data generation where the UIS has or is

supporting the implementation of AMPL. Countries that meet these criteria are Kenya (completed), Pakistan (currently administering) and Nepal (currently piloting).

- **Learning data:** We wanted a mix of countries who have already reported against SDG 4.1.1 (including countries that have reported regularly) and countries that have never reported. We want to include countries where the PAL Network is active, that has comparable learning data and has reported regularly (Kenya, Mexico) and observe these countries after PAL Network data collection is completed, so assess uptake of the data and value-add.
- **FCDO bilateral presence:** We wanted at least one country that does not have an active FCDO bilateral ODA programme (Senegal and Mexico).
- **Colonial heritage and regions:** We wanted at least one country with a Hispanic heritage (Mexico), and if possible, one with a Francophone heritage (Senegal), while also including countries from Latin America and the Caribbean (LAC) region (Mexico) and Asia (Nepal).
- **EiEPC:** We wanted to include at least two significant EiEPC countries (Somalia, Kenya).

In-country and desk data collection in 2025

30. Out of these six countries, in-person data collection took place in Ghana and Kenya. In Ghana, in-person data collection in 2025 was expected to allow collection of data on the results of ACER UK’s support for the 2024 national assessments and capacity building for the next round of assessments, alongside ALMA support of the same counterparts for the next round of assessments and implementation of the Accelerator support. For Kenya, the hope was that it would offer the opportunity to observe the PAL Network global training of partners in practice, an opportunity to engage with all partners of the Network in person. Table 12 below provides more detail on the individual country cases selected, including key context factors. Kenya also covered the contracted work of the NFER, with the KNEC to establish the EARC.

Table 12 Country cases

Country & FCDO Advisor presence	D4FL partners	Reporting against SDG 4	Change in learning outcomes	Region Language	EiEPC \$m and share of total humanitarian funding ¹⁰³	Why inclusion or swap out
Carried over from formative evaluation						
Ghana Education Advisor 2025 only	ACER UK, ALMA, Accelerator	No data on SDG indicator 4.1.1	N/A	West Africa, English	No: \$0.006m; 0.2%	West Africa, overlap of three partners No reporting to date By end 2025 one round of ACER UK and ALMA will be done, and Accelerator would have made progress WWHGE
Kenya Education Advisor 2025 and 2026	PAL (NFER partner based in Kenya)	Reported primary maths and reading against SDG 4.1.1 (2019,2021,2023)	Large decrease in foundational maths and reading Yes	East Africa, English	Yes: \$6.4m, 2.7% (inc FCDO)	East Africa, PAL Network on its own w.r.t. national government, and a country that has learning data and was part of UIS pilots. Plus, location of

¹⁰³ UNOCHA, 2024

Country & FCDO Advisor presence	D4FL partners	Reporting against SDG 4	Change in learning outcomes	Region Language	EiEPC \$m and share of total humanitarian funding ¹⁰³	Why inclusion or swap out
						NFER regional programme. WWHGE country UIS-supported AMPL completed.
Nepal Education Advisor 2025, with 2026 to be decided	ALMA, PAL Network	No data on SDG indicator 4.1.1	N/A	South Asia, English	No: \$0.69m 2.2%	Overlap of ALMA and PAL, located in East, follow-up in 2025 to observe progress, but can be left out for 2026. Possible SCALE country Piloting UIS-supported AMPL
Senegal No Education Advisor 2025 and 2026	PAL Accelerator	Reported primary maths and reading against SDG 4.1.1 (2014, 2016, 2019) + later grades + early grade	Increase in foundation al reading but small decrease in foundation al maths	West Africa, French	No: \$0.3m 5%	West Africa, Francophone, no FCDO bilateral programme, Accelerator is implementing assessments, working with EGRA model.
Somalia Education Advisor 2025 and 2026	ACER UK ALMA	No data on SDG indicator 4.1.1	N/A	East Africa, English	Yes: \$30m 4.3% (inc by FCDO)	EiEPC country, ACER UK and ALMA overlap, no data. SCALE country
New						
Mexico No Education Advisor 2025 and 2026	PAL	Reported primary maths and reading against SDG 4.1.1 (2013, 2019) + later grades + early grade	Decrease in foundation al maths and small decrease in foundation al reading	LAC, Spanish	No: \$1.7m, 15%	Latin America, no FCDO bilateral programme, PAL operating where some data and reporting, but not a UIS pilot country Can swapped with Columbia for 2025 and 2026, but Mexico is more clearly making progress.

Undertaking country cases

31. For each remote data collection country, we had 5 days of core team, 4 days of national consultant and 2.5 days of researcher time. We used this time to build a country dossier, conduct interviews with the country partner teams, country counterparts including the education/assessment authorities, and FCDO education advisors, and write up the interviews, and complete the internal country case study data analysis matrix, drawing conclusions on lessons for the evaluation report. Table 13 provides details on the KIIs and FGDs conducted.

32. For each in-country fieldwork, we had 6 days of core team and 7 days of national consultant time. We used this time to update the country dossier, set up the country fieldwork, conduct interviews with the country partner teams, country counterparts (national education authorities and assessment authorities) and other country stakeholders like teacher professional bodies or EiEPC stakeholders, and conduct workshops with country stakeholders as relevant. The core-team member assigned to the country spent a week in-country with the national consultant. The team reserved one day of the five days in country to jointly complete a draft case study data analysis matrix, and provided an on-line presentation to country stakeholders to validate the emerging

findings. We aimed to schedule visits to align with key support activities, so that we could observe these too, but was not able to achieve this. We sequenced activities so that we could pick up on emerging themes and include workshops to explore, clarify and validate findings. Table 13 provides details on the KIIs and FGDs conducted.

Case study reports

33. The evaluation did not produce country case reports. It recorded evidence from each case in a country case evaluation analytical matrix; these remained internal to the team. They fed into the consolidated analytical framework. Conclusions and recommendations were included at the programme level.

Summary of key informant interview and workshop sampling

34. Table 13 summarises our respondent sampling approach and anticipated and achieved sample sizes for each primary data collection activity.

Table 13 MTE respondent selection for data collection

Activity	Intended sampling strategy	Intended n.	Achieved n.
Global level			
Interviews with members of the GCFL (Non-D4FL global assessment actors)	Three members interviewed	3	3
ToC workshops with partners	ACER UK and the PAL Network declined a ToC workshop to start the MTE. We conducted one with the World Bank, to discuss the explicated WB ToC, and with UNICEF and NFER to develop a ToC.	5	3
Partner interviews	1-2 interviews with strategic and operational team members per organisation (ACER UK, PAL, UIS, GEM-R, World Bank, NFER and UNICEF).	7-10	7
Workshop on assessment in the classroom and D4FL	1 workshop with the D4FL partners (attended by ACER UK, World Bank ALMA and Accelerator, NFER, PAL Network, UNICEF representatives)	1	1
Workshop with FCDO education CMP leads	1 workshop with senior staff from the WWHE, SCALE, Ed Tech Hub, D4FL, and Team Leader for the Education Programme Team	1 ¹⁰⁴	1
Learning and reflection workshops	1 Workshop each with the World Bank, ACER UK and the PAL Network, 1 with the FCDO, and 1 joint workshop with FCDO and all partners. The FCDO D4FL Steering Committee attended the FCDO workshop	5	5

¹⁰⁴ In discussion with the FCDO D4FL Steering Committee, the team shifted from the individual interviews planned, to a workshop, for a more coherent discussion on cross-programme complementarity.

Activity	Intended sampling strategy	Intended n.	Achieved n.
	The global leads and senior team members from each partner attended the validation workshops. The joined-up workshop were attended by representatives of all the D4FL partners, including UIS and GEM-R.		
FCDO D4FL lead interview (1 update meeting at start of data collection; 1 interview mid-way; and 1 extended meeting to discuss conclusions and recommendations).		2	1 Formal KII 2 Meetings
Country-level (six sampled countries)			
Government interviews	At least one interview with Government representative(s) (or key in-country stakeholder(s) per sampled country. Ghana (5) Kenya (2) Nepal (2) Mexico (2) Senegal (3) Somalia (2)	6	16
Partner interviews	One interview with ACER UK/PAL network/World Bank/NFER country lead(s) per sampled country. Ghana (1 FGD with 3 participants, 1 KII), Kenya (3) Mexico (2) Nepal (2) Senegal (1) Somalia (2)	11	14
FCDO adviser interviews	One interview per sampled country where FCDO adviser in post. We did not interview the advisor in Ghana who was on maternity leave at the time of the fieldwork.	4	3
Other stakeholders	Interviews with stakeholders in assessment at country level. These were intended only for the in-person country cases, but were also done in Mexico and Nepal. Ghana (2) Kenya (0) Nepal (2) Mexico (1) Senegal (1) Somalia (0)	2	6
Totals	Interviews Workshops	35-38 12	56 10

Survey(s)

35. All countries,¹⁰⁵ including countries not covered in the country cases, were invited to participate in an online survey, in the MTE, via the D4FL partners and FCDO advisors. Country education/assessment authorities, partner country teams, and other country stakeholders were invited to complete the survey. Survey results are provided in Annex 9.

36. The survey was designed in August, tested early September and run in September/October. It therefore followed country fieldwork, and was intended as a rapid validation to test whether emerging themes against the evaluation framework apply in a wider number of countries. Only limited resources had been assigned to the survey, but it was nonetheless an important

¹⁰⁵ At present, the FCDO anticipate around 20-30 countries being supported by the D4FL programme. As the provision is demand-led, the final number could be higher or lower.

contribution, given the limited number of country cases. The survey questions were primarily closed-ended, using binary or ordinal Lickert-style questions to categorise and quantify respondents' opinions on emerging themes from the country cases, for example where different cases indicate divergent findings. Open-ended questions were used to provide opportunities for respondents to explain their answers.

37. We anticipated respondents from countries not supported by the D4FL programme also to be surveyed to avoid bias from positive selection of survey D4FL countries only, sharing some questions (e.g. on the D4FL programme's relevance to countries' needs) with respondents from countries where the D4FL programme is implemented. Target survey respondents included education officials and other country partners of the implementing partners for D4FL countries, and education officials for non-D4FL countries. The ET worked with the FCDO and implementing partners, as well as global stakeholders, to distribute the survey to respondents.

Analysis

38. We took a similar approach to data management and analysis as outlined for the formative evaluation.¹⁰⁶ Key elements were:

- We analysed survey responses descriptively using unweighted frequencies and 'cross-tab' analysis. Where sample sizes allowed, we used cross-tabulation to explore differences between groups (for example countries working with different partners or in different regions) and to investigate relationships between different key survey variables – for example challenges experienced and views on the value of the support.
- We utilised secondary analysis of quantitative data sources (see for example under Finding 1 and Finding 2). This analysis is mostly descriptive, compiling data sources and presenting key figures and trends (the numbers of countries reporting internationally, number of countries assessing foundational learning outcomes at 2nd and 3rd grade and at the end of primary, the number of countries who have nationally representative data at those levels and the number of countries who can report trends at each level). Where relevant, we had intended to review any assessment data collected and analyse it with D4FL support, for example for extent of disaggregation, however, in the end there were no data available for review/analysis, as in none of the cases data had been reported yet.
- As for the formative evaluation, analysis continued throughout using the Microsoft excel-based framework matrix (coding data from sources – in rows – against the evaluation questions and assessment criteria – in columns) so that we could explore emerging findings in later data collection activities. This closed coding was then analysed through open-ended grounded coding to explore meta-themes. MAXQDA was used by some team members for this qualitative analysis.
- We triangulated findings from different data sources to address the evaluation questions, synthesising evidence first within country cases and then across countries and at a global level. We used the emerging findings matrix provided in Annex 6 of the IR to summarise findings and evaluative judgements, synthesise the evidence, and assess the weight of that supporting evidence.

¹⁰⁶ Analysis was predominantly done by the team and AI (Microsoft CoPilot) was only used to do a preliminary coding of long reports and for a few interview transcripts, to supplement interview notes, in response to questions formulated by the team. The initial AI outputs were reviewed before use, to ensure accuracy and consistency with other data.

- We did not undertake full contribution analysis in this round of the evaluation, because the findings on results and impacts are still significantly based on surmise, given progress on D4FL implementation. We did take care during data collection to collect information on other factors and actors that could affect DFL results. We will utilise this information on D4FL and other actors' contributions, as well as context factors, to analyse the D4FL's contribution claims and evidence gaps at the start of the final summative evaluation, and collect additional evidence during this evaluation. The final summative evaluation report will provide the results of the contribution analysis on observed results against relevant questions, to set out the strength of D4FL's contributions, and substantiate the claim.

39. The team held two internal analytical workshops, one in June to review data collection to date, in particular the country cases, and extract themes for testing in the survey and subsequent data collection. The second workshop was held in October, and provided the headline findings which were then tested with the D4FL programme management and key partners.

Validation

40. Besides validating country case study findings during in-country fieldwork, we:
- Undertook validation and sense-making workshops with the FCDO and partners in October/November of 2025. We conducted one workshop with each partner (ACER UK, PAL Network, World Bank) and one workshop with the FCDO which were focused on validation of findings. The NFER and UNICEF was covered through final interviews. This was followed by a joint workshop with FCDO and all partners (ACER UK, PAL Network, World Bank, UNICEF, NFER, GEM-R) to reflect on and make sense of the findings for recommendations. A final discussion was held with the D4FL programme management, to present the conclusions and discuss recommendations, before finalising the draft report.
 - Provided a first draft report to the FCDO for comment, and a second draft report to partners at the same time as the EQUALS review. Comments on the report from partners used an excel-based comment matrix, which provides for feedback from the team on how comments were incorporated in the final draft.

Approach to equity

41. We mainstreamed equity considerations throughout the design, implementation and analysis of this evaluation.

42. In designing the EQs we considered the equity dimensions of the assessment and of the D4FL programme. This is explicit in some EQs – for example sub-question 1.4 addresses the “equity and inclusion objectives” of the D4FL programme, 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, we 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?

43. It was not possible to meet with organisations representing different marginalised groups during our in-person country fieldwork, as intended. In both cases interviews could not be secured

in the week in-country. We will review the commitment to meet with such representatives when considering lessons from this evaluation round, for drafting the plan for the final summative evaluation in January 2026.

44. 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

45. 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 data collection and the evaluation results do no harm to participants or their communities.

46. 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 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 review 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 FCDO to ensure that dissemination of findings is adequate, appropriate and well balanced.

Team Roles and Responsibilities

47. There is no known conflict of interest concerning any team members. All core team members are 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 MTE report. The deputy team leader provided strong support to the team leader, taking the lead on selected processes.

48. 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 Deputy Team Leader</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 • 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
<p>Christine Fenning Evaluator</p>	<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 her areas of responsibility
<p>Javier Pereira Evaluator</p>	<ul style="list-style-type: none"> • Provided evaluation and development financial flows expertise • Participated in data collection phase • Conducted relevant and selected interviews • Main responsibility for design and implementation of survey • Participated in the team workshop to draw findings and conclusions together from the body of evidence collected during the evaluation • Inputted into the ER in line with his areas of responsibility
<p>Additional Experts</p>	
<p>Principal expert Rachel Outhred</p>	<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

Team member and role	Responsibilities/coverage
	<ul style="list-style-type: none"> Provides inputs into ER, ensuring quality standards and agreed timelines are met
Technical expert 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 Ellen Smith and Sara Bashir Malik 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
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 7 Mid-Term Evaluation Framework

MID-TERM EVALUATION FRAMEWORK	
CONCLUDING / OVERARCHING QUESTIONS	<p>Is the D4FL programme producing the intended results?"</p> <p>Is the D4FL programme representing good value for money?</p> <p>What changes should the FCDO and implementing partners make, to increase the likelihood of programme success in its final year?</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
CRITERION: RELEVANCE			
The extent to which D4FL's objectives and design respond to the needs of LIC/LMIC country-level and global policy makers on and stakeholders in education outcome monitoring and standardised learning assessments.			
1. To what extent does D4FL address the needs of LICs/LMICs?	1.1 To what extent are the programme impacts and objectives relevant to the needs of LICs/LMICs?	<p>Improved foundational learning outcomes is a priority need of LICs/LMICs.</p> <p>Generating comparable and disaggregate learning data to report globally on learning progress is a priority need of LICs/LMICs.</p> <p>Generating comparable learning data that can be used in decision-making and the design and monitoring of learning interventions is a priority need of LICs/LMICs.</p> <p>Strengthening capabilities to use learning data is a priority need of LICs/LMICs.</p> <p>Governments and other in-country stakeholders are interested in working with D4FL to address their learning data needs.</p>	<p>Interviews with global learning data experts</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p> <p>Surveys of LIC/LMIC governments</p>
	1.2 To what extent are programme activities relevant and aligned to the needs and contexts of LICs/LMICs'?	<p>The D4FLs choice of partners are relevant to the needs and contexts of LICs/LMICs.</p> <p>Implementing partners have appropriate processes in place to assess in-country stakeholders' needs and the context, and tailor their support, including when/if circumstances change.</p> <p>D4FL-funded activities respond to in-country stakeholders' needs (including adapting to changing needs) and are aligned to context, including country policies, plans,</p>	<p>Document review of partners' plans and progress reports</p> <p>Review of country primary and secondary documents</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
		<p>education structures, logistical features (e.g. timing and mode) of assessment activities and capacities.</p> <p>D4FL supports the right functions within countries' assessment systems to produce learning data to meet global and in-country needs (systems include assessment for learning, examinations, citizen-led assessments, sample-based standardised assessment).</p>	<p>Country-level observation of activities</p> <p>Global-level interviews with partners and FCDO lead</p> <p>Survey of participating governments</p>
<p>2. To what extent do D4FL's objectives and activities address the needs of the global community?</p>	<p>No sub-questions</p>	<p>Improved foundational learning outcomes remains a global priority.</p> <p>There is a continued global commitment to filling global learning data gaps to monitor education quality and learning outcomes.</p> <p>D4FL's objectives and activities respond to the most urgent global data learning gaps, and address the most significant global barriers to LICs/LMICs generating, reporting and using learning data.</p> <p>The programme is able to adapt when/if the needs of the global education sector change.</p>	<p>Document review to understand extent to which countries are currently reporting internationally against SDG 4.1</p> <p>Interviews with external assessment experts</p> <p>Interviews with global learning assessment actors</p> <p>Global-level interviews with partners and FCDO lead</p> <p>Document review of literature on global assessment landscape</p>
<p>CRITERION: COHERENCE</p> <p>The extent to which D4FL's objectives, design and activities are internally coherent and complementary, and complementary to other learning assessment initiatives.</p>			
<p>3. To what extent is D4FL a coherent programme across strands and partners?</p>	<p>No sub-questions</p>	<p>FCDO, implementing partners and countries are working towards common objectives, in particular the D4FL programme objectives, namely filling global data gaps, the availability of more disaggregated data that can be used, and increased country capability to use learning data to design and monitor foundational learning interventions.</p> <p>FCDO and implementing partners present a coherent narrative about the value and complementarity of different types of data.</p>	<p>Document review of D4FL documentation, including the business case and annual reports</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p> <p>Global-level interviews with partners and FCDO lead</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
		FCDO and implementing partners take appropriate and effective steps to link up and ensure their work is complementary and not duplicating at global and country level.	Survey of participating governments
4. To what extent are D4FL activities coherent and complementary with other global learning data efforts?	No sub-questions	<p>D4FL is adding value to the international learning data landscape while not duplicating efforts.</p> <p>D4FL is not duplicating other international learning assessment initiatives at country level.</p> <p>D4FL contributes to increased harmonisation and coordination across global efforts to improve the tracking, generation and use of foundational learning data.</p>	<p>Global-level interviews with partners and FCDO lead</p> <p>Interviews with global learning assessment actors and experts</p> <p>Partners' D4FL plans and reports</p> <p>Country level interviews with partners, government, other stakeholders and FCDO advisors</p> <p>Global-level interviews with the FCDO lead and leads of other central programmes</p> <p>Survey of participating countries</p>
5. To what extent is D4FL coherent and complementary with other FCDO programmes?	No sub-questions	<p>D4FL and other centrally managed FCDO programmes (CMPs), in particular the What Works Hub for Global Education (WWHGE) and Scaling Access and Learning in Education (SCALE) are cooperating at the global level to strengthen learning outcomes in LICs/LMICs.</p> <p>Activities in the same country and funded by D4FL and other centrally managed FCDO programmes, in particular the WWGE and SCALE,¹⁰⁷ are well-coordinated and complementary towards improved learning outcomes.</p> <p>D4FL activities are supported by and support bilateral aid programmes to achieve D4FL and FCDO objectives.</p>	<p>Document review of FCDO D4FL and other CMPs' documentation,</p> <p>Review of D4FL partner reports</p> <p>Global-level partner interviews</p> <p>Global interviews with FCDO lead and leads of other FCDO CMPs</p> <p>Country-level interviews with partners of D4FL and other CMPs, FCDO education advisors/teams, government and other country stakeholders</p>

¹⁰⁷ Other relevant CMPs are FCDO support for the Ed Tech Hub, and the Education in Emergencies Programme.

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
<p>CRITERION: EFFECTIVENESS The extent to which D4FL has and is likely to achieve intended outcomes.</p>			
<p>6. To what extent is D4FL leading to the availability of more disaggregated learning data in LICs/MICs, that can be used to design and monitor foundational learning interventions?</p>	<p>6.1 To what extent is D4FL increasing the availability, quality, and/or comparability of foundational learning data in LICs/MICs?</p>	<p>D4FL-funded activities have led (or are likely to lead) to more data on foundational learning being available in supported countries.</p> <p>D4FL activities have improved (or are likely to improve) the quality of learning data available in supported countries in one or more of the following dimensions:</p> <ul style="list-style-type: none"> • Frequency or regularity of available data • Number of grades of primary education that are assessed • Number of learning domains that are assessed (literacy, maths, other aspects of learning) • Data completeness, i.e. Increased or better population representativeness¹⁰⁸ • Data accuracy, i.e. the data is likely to be a reliable reflection of reality • Improvement in perceptions of data credibility and trustworthiness. <p>D4FL activities have improved (or are likely to improve) the comparability (temporally, spatially) of learning data in supported countries.</p>	<p>Document review of partner plans and reports</p> <p>Document review of partner deliverables at country level</p> <p>Interviews and workshops with D4FL partners at global level</p> <p>Interviews with the D4FL lead and workshops with FCDO</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p> <p>Survey of participating countries</p>
	<p>6.2 To what extent is D4FL leading to more and better learning data for children affected by drivers of exclusion from learning?</p>	<p>D4FL activities have led (or are likely to lead) to greater inclusion in learning assessments of one or more of the following often-excluded population groups in supported countries:</p> <ol style="list-style-type: none"> 1. Girls/Boys (depending on context) 2. Children with disabilities 3. Children affected by emergencies and protracted crises, especially refugee children and internally displaced children 4. Out-of-school children. <p>D4FL activities have improved (or are likely to improve) availability of data that can be used to compare learning gaps between one or more of these often-excluded population groups and their peers in supported countries.</p>	

¹⁰⁸ Note that progress on disaggregation by population groups (other than by location) is assessed under the next sub-question.

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	<p>6.3 To what extent is D4FL leading to learning data that can be and is useful for application and decision-making in-country?</p>	<p>D4FL activities have led (or are likely to lead) to increase in the availability of data that:</p> <ol style="list-style-type: none"> 1. addresses issues identified by national stakeholders as strategically important 2. addresses the specific learning needs and gaps within the setting 3. can respond to requests from national policy makers 4. is easily understood by decision makers and other relevant actors 5. is available in the right format at the right time 6. can inform decision making on education resourcing, goals or foci 7. can be used to advocate for action on foundational learning. 	
<p>7. To what extent is D4FL leading to better capability in priority countries to use learning data to design and monitor foundational learning interventions?</p>	<p>7.1 To what extent is D4FL leading to increased capability in supported countries to use learning data in the design and monitoring of foundational learning interventions?</p>	<p>D4FL activities have built or (are likely to build) the capability of education authorities in supported countries to effectively use learning data through one or more of the following pathways,¹⁰⁹</p> <ol style="list-style-type: none"> 1. Building sufficient ownership of learning data 2. Building sufficient knowledge of the assessment design, and sufficient knowledge of the assessment's strengths and limitations 3. Building sufficient in-country capacity to analyse assessment data to answer decision-makers' questions 4. Building the appetite of education authorities to use learning data and its analysis <p>D4FL activities have built (or are likely to build) the capability of non-state actors in supported countries to effectively use learning data.</p>	<p>Document review of partner plans and reports, and partner deliverables at country/global level</p> <p>Interviews and workshops with D4FL partners at global level</p> <p>Interviews with D4FL lead and workshops with the FCDO</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p> <p>Country observation of activities</p> <p>Survey of participating countries</p>
	<p>7.2 To what extent do policies, targets and interventions in supported countries use or intend to use data generated through D4FL activities to</p>	<p>D4FL activities have led (or are likely to lead) to country authorities at different levels acting on learning, including,</p> <ol style="list-style-type: none"> 1. Using learning data to design, implement and monitor foundational learning interventions (such as strengthened curricula, teacher development, pedagogy and teaching materials) 2. Increased budgets for foundational learning improvement, and foundational learning, in national budgets 	

¹⁰⁹ Criteria for this question would be influenced by D4FL response to the recommendation in the formative evaluation, on how to incorporate D4FL ambitions on assessments in classroom in the TOC.

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	strengthen action on foundational learning?	3. Increased focus on foundational learning in country, in policies and other documents.	
8. To what extent and how is D4FL filling global learning data gaps?	8.1 To what extent is D4FL leading to an increase in the number of supported countries reporting learning data globally?	<p>D4FL-funded actions have led (or are likely to lead) to supported countries reporting, or reporting more comprehensively, against SDG4.1.1a or SDG4.1.1b.</p> <p>D4FL-funded activities have led (or are likely to lead) to supported countries reporting data that is recognised as internationally comparable by key assessment and education actors.</p>	<p>Document review of partner plans and reports, partner deliverables at country and global level</p> <p>Interviews and workshops with D4FL partners at global level</p> <p>Interviews with FCDO lead and workshops with FCDO</p> <p>Document review of case country policies, assessment frameworks, reports as applicable</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p> <p>Survey of participating countries</p>
	8.2 To what extent are the D4FL-supporting the development and implementation of methodologies that can track foundational learning for all, and fill global learning data gaps?	<p>D4FL is supporting the development of new methodologies to generate comparable, disaggregated data.</p> <p>D4FL-funded actions are demonstrating cost-effective and replicable ways of generating comparable and usable learning data, that can be taken up or scaled in practice in most LICs/LMICs (including countries affected by emergencies and fragile states), given assessment expertise and resources.</p> <p>D4FL-supported global public goods (or other outputs) on the generation, analysis, presentation and, use of learning data for education decision-making and the design and monitoring of foundational learning interventions:</p> <p>1. are/will be widely disseminated by the FCDO, D4FL partners and other actors</p>	<p>Document review of partner plans and reports, partner deliverables at country and global level</p> <p>D4FL financial data and complementary country-level financial data on costs where available</p> <p>Interviews and workshops with D4FL partners at global level</p> <p>Interviews/FGDs with non-D4FL global learning assessment actors</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
		<p>2. are/can be used by non-D4FL supported global education actors and/or non-D4FL LIC/LMIC country authorities.</p> <p>D4FL-supported models of TA are/can be used by other assessment actors to build the capacity of education authorities in LICs/LMICs to generate and use learning data for decision making and the design and monitoring of foundational learning interventions.</p>	<p>Interviews with FCDO lead and workshops with the FCDO</p> <p>Interviews with other FCDO leads</p> <p>Interviews with independent assessment experts</p> <p>Survey of non-D4FL countries</p>
<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>			
<p>9. How well is D4FL allocating resources to maximise the chance that outcomes were achieved?</p>	<p>9.1 To what extent has the FCDO allocated resources or is likely to allocate remaining resources to maximise the chance that outcomes are achieved?</p>	<p>The D4FL programme has sufficient resources (monetary and non-monetary) relative to its scope and methods to make a difference.</p> <p>FCDO has selected the right combination of implementing partners and allocated monetary resources between them optimally to maximise the chance that outcomes will be achieved.</p> <p>D4FL resources leverage other resources of partners.</p> <p>D4FL contracts include appropriate processes to negotiate/influence and monitor how D4FL implementing partners use resources to maximise the chance that outcomes are achieved.</p>	<p>Document review of D4FL Annual Report, partner plans and reports, and partner deliverables at country and global level</p> <p>Analysis of D4FL and partner-level financial data insofar available</p> <p>Interviews/FGDs with non-D4FL supported global learning assessment actors</p> <p>Interviews and workshops with D4FL partners at global level</p>
	<p>9.2 To what extent is D4FL implementing partners allocating resources to maximise the chance that outcomes are achieved?</p>	<p>Implementing partners' planned use and use of funding is only for activities that can be expected to contribute to D4FL targeted outcomes.</p> <p>Implementing partners have allocated resources and have processes in to plan activities in the right combination and sequence to maximise the chances that D4FL outcomes are achieved.</p>	<p>Interview with D4FL lead and workshops with the FCDO</p> <p>Country-level interviews with government, stakeholders, partners and FCDO advisers</p> <p>Survey of participating countries</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
<p>10. To what extent are partners delivering their activities in an efficient way?</p>	<p>10.1 To what extent are D4FL implementing partners using resources in a <u>timely</u> way?</p>	<p>Implementing partners have delivered activities to date in a timely way. FCDO and implementing partner processes are likely to facilitate the timely use of resources and delivery of activities. <i>Memo: Timely means within planned timeframes where applicable, and within timeframes that align with D4FL and country priorities for support and are acceptable to countries.</i></p>	
	<p>10.2 To what extent are D4FL implementing partners using resources <u>efficiently</u>?</p>	<p>Implementing partners have used resources efficiently to date. Implementing partners have and are prioritising the efficient use of resources. D4FL and implementing partners have processes or protocols in place to ensure that resources are used efficiently. <i>Memo: Efficient means within planned use of resources where applicable, to ensure the delivery of outputs at appropriate levels of quality at minimum cost, balancing and sequencing staff and other inputs to ensure that resources are not wasted, and avoiding the duplication of efforts.</i></p>	
<p>CRITERION: IMPACT The extent to which achieved D4FL outcomes are likely to contribute to significant progress against SDG4.1.1 and learning poverty reduction in developing countries.</p>			
<p>11. What evidence is there that D4FL's results will lead to improved quality of education and reduced learning poverty?</p>	<p>No sub-questions</p>	<p>Data generated through the D4FL programme is used to apply pressure at global and country level for action on foundational learning. Commitments by global actors and LIC/LMIC education authorities to improve foundational learning can be credibly linked to the D4FL programme activities and results. Commitments by global actors and LIC/LMIC education authorities to include or better serve populations often excluded from foundational learning can be credibly linked to the D4FL programme activities and results. Actions in supported countries observed against indicator 7.2 can be expected to improve the quality of education and therefore foundational learning outcomes.</p>	<p>Document review of partner D4FL plans, reports and published materials at global level Document review of global documentation in the field Interviews/FGDs with learning assessment experts Interviews/FGDs with non-D4FL supported global learning assessment actors Interviews / workshops with D4FL partners at global level</p>

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
			Interviews with D4FL lead and workshops with the FCDO Interview of other FCDO programme leads Country-level interviews with government, FCDO education advisors and other stakeholders Survey of participating countries Survey of non-D4FL LICs/LMICs
CRITERION: SUSTAINABILITY The extent to which the net benefits of the D4FL continue or are likely to continue			
12. Are in-country benefits likely to be sustained?	12.1 Are in-country changes in structures, policies and practice likely to be sustained?	Governments/national stakeholders in supported countries have ownership of D4FL-supported activities. Country capacities for assessment built through D4FL activities are sufficient and are likely to be sustained to ensure regular future assessments. Country assessment frameworks, policies and targets that institutionalise learning assessment, are not likely to be overturned. Government and/or civil society in countries supported by D4FL have committed to make financial, human, technical resources available or seek resources to repeat assessments regularly. D4FL-supported country authorities who have been supported to report against SDG 4.1.1a and 4.1.1b are committed to doing so regularly.	Document review of partner plans and reports Interviews / workshops with D4FL partners at global level Interviews with D4FL lead Review of country documentation Country-level interviews with government, stakeholders, partners and FCDO advisers Survey of participating countries
	12.2 To what extent has D4FL generated more demand in supported countries for foundational	D4FL-funded foundational learning interventions have led (or are likely to lead) to demand from education authorities in supported countries for learning data to design and monitor these and other learning interventions, or for decision making. D4FL-funded interventions have led (or are likely to lead) to demand by non-state actors in supported countries for learning data to monitor foundational learning interventions, and for action on learning improvement.	Document review of partner plans and reports Interviews / workshops with D4FL partners at global level Interviews with D4FL lead

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	learning data?		Review of country documentation Country-level interviews with government, stakeholders, partners and FCDO advisers Survey of participating countries
13. What long-term changes to the global assessment landscape are likely to result from D4FL?	13.1 To what extent has D4FL contributed to clear, consistent and useful requirements for international reporting?	D4FL activities have contributed to the UIS criteria being stable so that methodologies and country capabilities developed under D4FL to report data globally, remain relevant for this purpose. D4FL activities have contributed to the UIS criteria being clear and useful to national governments. D4FL activities have contributed to the development of criteria for global comparability that are achievable by most LICs/LMICs.	Document review of global documentation in the field Interviews with learning assessment experts Interviews/FGDs with non-D4FL supported global learning assessment actors Country-level interviews with government, stakeholders, partners and FCDO advisers Interviews / workshops with D4FL partners at global level Interviews / workshops of D4FL management and other FCDO programmes Survey of non-D4FL countries
	13.2 To what extent has D4FL contributed to methodologies that are recognised by global actors as	Global learning assessment actors (other than UNESCO) and learning experts believe D4FL-supported methodologies should be recognised as valid to track learning in LICs/LMICs. Data generated by new D4FL-supported methodologies are used by reputable global institutions / researchers to analyse tracking progress on learning in LICs/LMICs.	Document review of global documentation in the field Interviews with learning assessment experts Interviews/FGDs with non-D4FL supported global learning assessment actors

MAIN QUESTION	SUB-QUESTIONS	ASSESSMENT CRITERIA	DATA SOURCES
	<p>valid to track learning in LICs/LMICs?</p>	<p>New methodologies developed with support of D4FL have been (or are likely to be) recognised by UNESCO as valid to generate or report globally comparable learning data.</p>	<p>Interviews with FCDO lead and workshops with the FCDO Interviews with other FCDO programme leads</p>
	<p>13.3 To what extent has D4FL increased demand for foundational learning data outside of supported countries?</p>	<p>More LICs/LMICs approach UIS, FCDO and other implementing partners and assessment stakeholders for support with learning assessments and/or international reporting after seeing the outputs of D4FL. D4FL partners are receiving requests for support with learning assessments and/or international reporting linked to their D4FL work. State and non-state stakeholders in non-D4FL LICs/LMICs want to generate and see it is as possible to generate regular foundational learning data.</p>	<p>Interviews with learning assessment experts Interviews/FGDs with non-D4FL supported global learning assessment actors Interviews / workshops with D4FL partners at global level Interviews with other FCDO programme leads Survey of non-D4FL countries</p>

Annex 8 Use and Influence Plan

1. 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 two parts:
 - The Use and Influence Plan, including planned dissemination outputs and engagement activities, providing a restatement of the principles, outputs, and engagement strategies as articulated in the IR.
 - 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

2. As described in the IR, the Use and Influence Plan adopted 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:

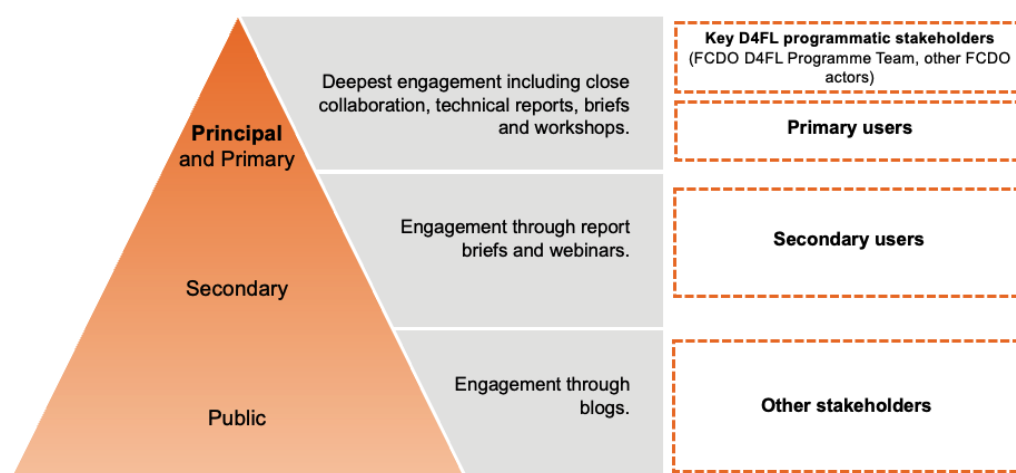
Principles for Evidence Uptake

3. Reflecting the variety of stakeholders and uses for the evidence from this study and of this report, we are deploying a variety of outputs and evidence uptake strategies. Across the range of outputs, we are employing the following principles:
 - **Variety of materials:** We take a pyramid approach to production of outputs (see Figure 13 overleaf). This recognises that different groups have different time availability to engage with research outputs. Longer written reports are supplemented by shorter summary materials, or mixed media to allow for engagement by a range of audiences.
 - **Accessibility:** The research outputs will be accessible to the range of audiences that can benefit from the research. While some outputs will need to include technical descriptions of findings, we aim to minimise the use of jargon and technical language and include simple descriptions of approaches and results where possible.
 - **Active engagement:** We understand the uptake of evidence as being an active process. We will work to support passive communication (e.g. written materials) with active communication (e.g. workshops) wherever possible, particularly for primary audiences.

Use and influence instruments

4. The evaluation will produce the following outputs: evaluation reports and evaluation digests for the formative and mid-term evaluations. For the summative evaluation, these two instruments will be supplemented by an evaluation brief, and an evaluation blog.
 - **Evaluation Digest:** A short summary of 2 to 4 pages, communicating the key findings, recommendations and conclusions.
 - **Evaluation Brief:** The brief presents the evaluation findings and recommendations in a plain English format, for greater uptake.
 - **Evaluation Findings Blog:** The Blog will summarise the findings for a non-technical audience and be submitted for publication to several sites where education actors frequent. Examples include UKFIET, the GEM Report website etc.

Figure 13 Pyramid approach to engagement



5. Active engagements include the following:

- Learning and reflection workshops with the D4FL Steering Committee and implementing partners (jointly and separately), as well as validation workshops with the implementing partners.
- In the summative evaluation, these workshops will be supplemented by presentation of the final evaluation findings to the D4FL Steering Committee, to other primary users, and a webinar for secondary users.

Table 14 Progress on use and influence plan

Stakeholder	Interest and Involvement in the Evaluation	Use and influence actions	Progress
PRINCIPAL USERS AND INFORMANTS: FCDO¹¹⁰			
D4FL Programme Team and the Girls Education Department of the FCDO	The D4FL Programme Team have a high interest and involvement in the evaluation. The team wrote the TORs for and commissioned the evaluation, and will manage the evaluation contract to ensure it meets the needs of FCDO and other evaluation audiences.	<p>The Evaluation will engage with the programme team and larger department through:</p> <p>Meetings to align the evaluation with the principal users' needs and priorities</p> <p>Written outputs, including the evaluation plan, preliminary findings presentations, conclusion and recommendation presentations</p> <p>Validation and learning and reflection workshops</p> <p>Presentations of the final summative evaluation to the D4FL steering committee.</p>	<p>For the MTE the evaluation team:</p> <ul style="list-style-type: none"> • Held meetings with the D4FL Senior Responsible Officer (SRO) throughout the MTE, in addition to the formal interviews, to ensure the relevance of the evaluation to the D4FL • Conducted workshops, including a workshop on coherence between FCDO centrally managed education programmes, the preliminary findings workshop and the

¹¹⁰ The IR presented the FCDO Girls Education Department as a primary rather than principal user of the evaluation. As the D4FL is a discrete programme ending in 2027, the Department will be the principal user of the evaluation for future programming. For this reason we believe it should also be a principal user.

Stakeholder	Interest and Involvement in the Evaluation	Use and influence actions	Progress
			<p>learning and reflection workshop</p> <ul style="list-style-type: none"> • Conducted follow-up meetings with the SRO to discuss recommendations • Completed the MTE evaluation plan, MTE report, a presentation on the MTE findings, conclusions and recommendations, and the Evaluation Digest
FCDO Education Advisers	While not directly involved in the evaluation, the evaluation results are of high interest to FCDO Education Advisers.	<p>The D4FL Senior Responsible Officer may distribute the formative and mid-term evaluation report to all advisors. The FCDO Education Advisers will receive the final summative evaluation brief</p> <p>The FCDO Education Advisers network will participate in the presentation of the final evaluation in a webinar.</p>	No action required in the MTE.
PRIMARY USERS and INFORMANTS			
Implementing partners, global learning assessment community, country governments			
UIS and GEM-R	<p>UIS is both a partner in the D4FL Programme, and a recipient of D4FL Programme outputs, as the custodian of data against SDG4.1.1</p> <p>GEM-R is a D4FL partner and user of its programmatic outputs.</p>	<p>UIS and GEM-R participate in learning and reflection workshops. The UIS and GEM-R validate the evaluation report as D4FL partners, and will receive the final report via the D4FL SRO</p> <p>The UIS and GEM-R will participate in the presentation of the final evaluation in a webinar.</p>	<p>Representatives from UIS and MTE participated in the joint learning and reflection workshop. UIS provided comments on the report which were incorporated, and will receive this final MTE report and Digest.</p>
Other D4FL partners	<p>The World Bank FLC Umbrella Trust Fund ALMA and Accelerator Programmes; ACER UK; the PAL Network; NFER and UNICEF global teams and country teams/partners</p>	<p>Partners' global representatives participate in validation and learning and reflection workshops. Partners validate the draft evaluation reports, and receive the final reports and Digests.</p> <p>Partners will participate in the presentation of the final evaluation in a webinar.</p> <p>Partners will be recipients of the final summative evaluation report brief</p>	<p>Validation workshops were conducted with the World Bank, ACER UK and the PAL Network, but not yet with UNICEF and NFER, given progress on their D4FL activities.</p> <p>Representatives from all partners attended the joint learning and reflection workshops.</p>

Stakeholder	Interest and Involvement in the Evaluation	Use and influence actions	Progress
			<p>The global teams of all partners commented on the MTE report and their comments were incorporated, and responded to in a response matrix.</p> <p>The partners will receive the final MTE evaluation report and digest.</p>
<p>Other stakeholders in the global learning assessment architecture</p>	<p>Other stakeholders in the global learning assessment architecture and peers in the Global Coalition for Foundational Learning, generally are informants for the evaluation, and would be interested in its results as stakeholders in the assessment landscape.</p>	<p>These stakeholders can access the formative and mid-term evaluation outputs from the FCDO website, as well as the final evaluation report and digest.</p> <p>These stakeholders will be recipients of the final summative evaluation brief.</p> <p>These stakeholders will participate in a webinar on the final summative evaluation findings, recommendations and conclusions</p>	<p>No action required in the MTE.</p>
<p>National Governments participating in the D4FL Programme</p>	<p>These stakeholders are informants in the country cases, and are interested in the evaluation results</p>	<p>Stakeholders are engaged through country case study processes. They can access the evaluation reports and digests on the FCDO website.</p> <p>These stakeholders will be an audience for the final summative evaluation brief.</p> <p>These stakeholders will participate in a webinar on the final summative evaluation findings, recommendations and conclusions</p>	<p>Stakeholders were engaged in the country cases, but only to a limited extent in the desk studies</p>
<p>SECONDARY USERS and INFORMANTS Other countries, other education actors</p>			
<p>Other National Governments, Broader education actors</p>	<p>National Governments not participating in the D4FL Programme will have a low level of interest and very little or no involvement in the evaluation. These governments might be interested in the lessons learnt.</p>	<p>These stakeholders can access the evaluation reports and Digests on the FCDO website.</p> <p>These stakeholders will be the audience for the evaluations findings blog</p> <p>These stakeholders will be invited to a secondary stakeholders’ webinar on the final summative evaluation findings, recommendations and conclusions.</p>	<p>No action required in the MTE</p>

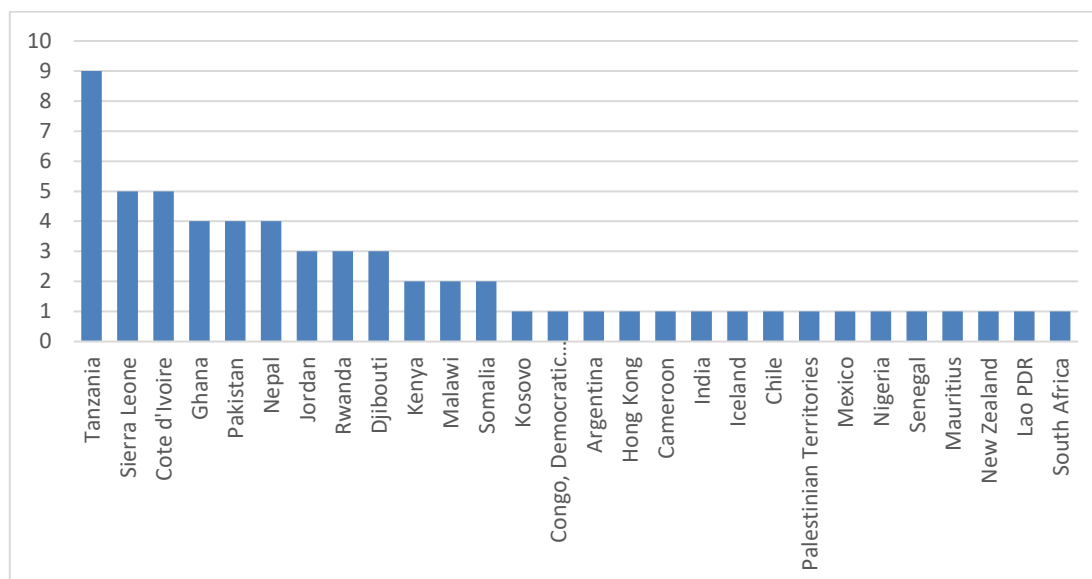
Annex 9 Survey Results

1. An e-survey was used to collect primary data from evaluation stakeholders. The survey was developed by the evaluation team and comprised two parts. The first part consisted of a set of questions seeking to understand perceptions about foundational learning at country level. This part had a strong focus on learning data. The second part comprised country-specific questions to collect stakeholders' perception on interventions supported by the D4FL programme. Only participants from these countries were prompted with the country-specific questions.
2. The survey was first tested by the evaluation team. Subsequently, it was tested with selected respondents from the target respondent groups, including the FCDO, global assessment stakeholders, partners and country government representatives. Questions were adjusted following comments received.
3. The survey was disseminated through multiple channels. FCDO supported dissemination at country level. D4FL partners also disseminated the survey through their networks, including country focal points and/or on social media. In addition, the survey was shared by the OECD Education and Skills directorate through its networks via newsletters, and at the ADEA Triennale on Education using a QR code.
4. The survey received a total of 63 replies. Given the dissemination approach, the total number of recipients of the invite to complete the survey is not known. As a result, it is not possible to estimate a response rate.
5. Open questions were reviewed one by one and categorised to identify themes. This process was refined through multiple steps. A first analysis was conducted with a sample of questions (20) and reviewed by a second team member. The approach was then applied to the full sample. The results were then reviewed by a second team member. The emerging set of themes was used for the final analysis.

Statistical questions

6. The survey received a total of 63 answers: 33 from female stakeholders (52 percent) and 30 from male stakeholders (48 percent). The survey received responses from 29 different countries. The average number of respondents per country was 2.2. In general, the number of respondents was higher in countries where D4FL partners were active. This was the result of the efforts made to disseminate the survey through the partners in order to capture perceptions on the activities supported by D4FL.

Figure 14 Survey respondents per country



7. Most respondents, worked for government (25 replies or 40 percent) and development partners (24 replies or 38%). Among those working for government, 92 percent (23 replies) worked for the education ministry. Among these, 10 worked in policy and planning units, 8 worked in units related to standards and examinations and the remaining 5 in units focusing on primary or pre-primary education.

Table 15 Survey respondents by affiliation

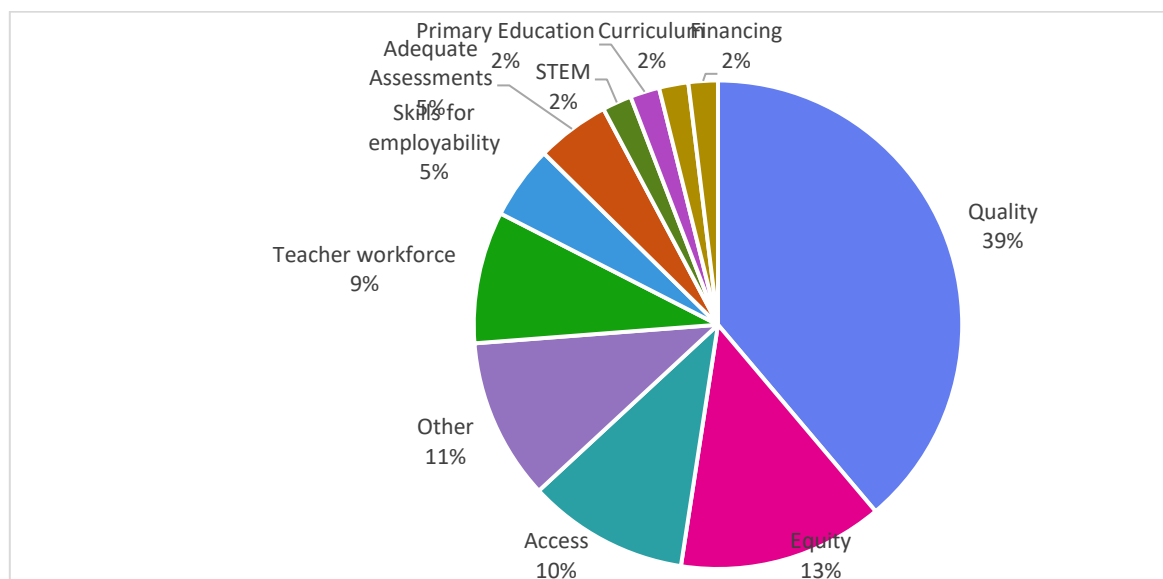
Question	
What sector do you work in?	
I work for a development partner institution	24
I work in the private sector	1
I work in school, college or university (e.g. a teacher or school leader)	4
I work for a civil society organisation (CSO or NGO)	9
I work in a government ministry, department or agency	25
What ministry, department or agency do you work for?	
Other	2
Education	23
Which of the following sub-sectors best matches the functions of the department or unit you work for?	
Policy and planning	10
Standards and examinations	8
Primary education	3
Pre-primary education	2

Common questions

Education priorities

8. The survey started by asking respondents about what they thoughts was the biggest education priority in their country. This was an open question. The responses (63) were categorised used a series of codes or themes. The results of the analysis are presented in the figure below.

Figure 15 Biggest education priorities



Level of government priority for different stages of education

9. Respondents were asked to rank different education stages based on what they think was a higher and lower priority for their government. The answers were translated into a score by assigning them a numerical value based on the ranking. Each answer that ranked a stage in first place was assigned a value of 4, while each answer ranking a stage in 4th place was assigned a value of 1.

Table 16 Perceived government priority per education stage

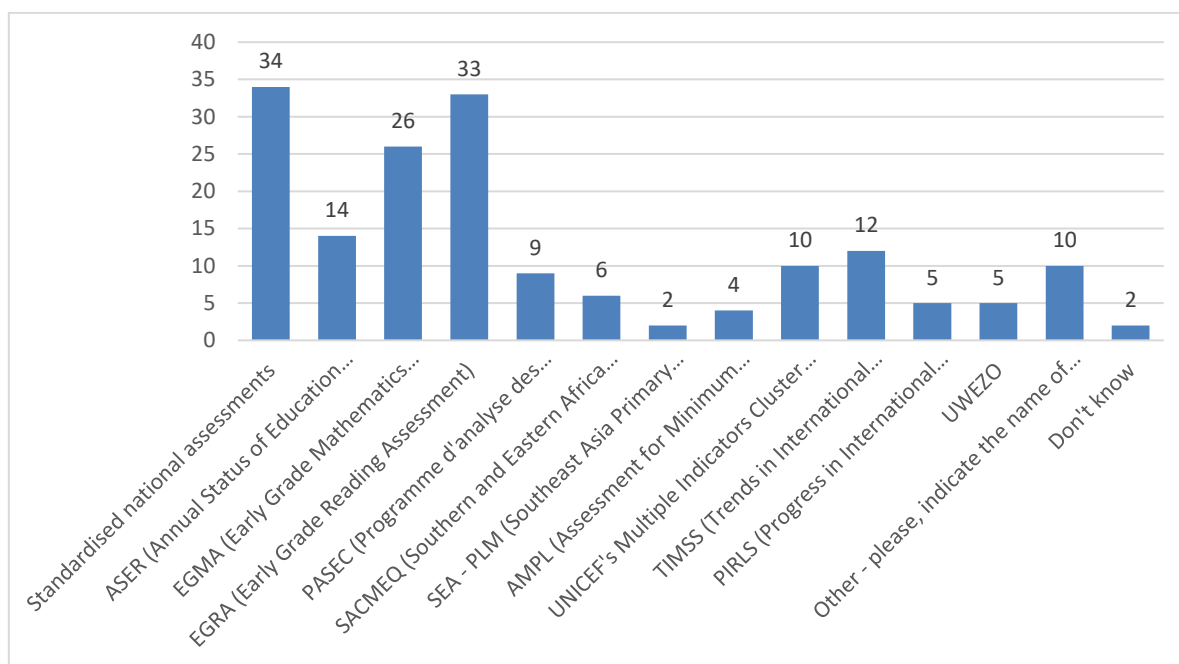
Ranking - all respondents	Pre-primary	Primary	Secondary	Post-secondary
1st	16	32	6	9
2nd	12	16	27	8
3rd	16	12	25	10
4th	19	3	5	36
Avg. value (1st=4; 4th=1)	2.40	3.22	2.54	1.84
Count	63	63	63	63
Government				
1st	9	11	2	3
2nd	4	9	9	3
3rd	8	4	12	1
4th	4	1	2	18
Avg. value	2.72	3.20	2.44	1.64
Count	25	25	25	25
Development partners				
1st	2	17	1	4
2nd	8	2	12	2
3rd	4	5	9	6
4th	10	0	2	12
Avg. value	2.08	3.50	2.50	1.92
Count	24	24	24	24

Data on learning outcomes

10. Respondents were asked whether their countries have nationally representative data on learning outcomes such as reading and math scores. Most respondents (54 respondents or 86 percent) confirmed this was the case. Out of these 54 respondents, 49 respondents (91 percent), stated that data covers the foundational learning stages.

11. Subsequently, respondents were asked to list the tools used to collect nationally representative data on learning outcomes such as reading and math scores.

Figure 16 Tools used to collect data on learning outcomes



Access to foundational learning

12. Respondents were asked to rank the importance for the government in their countries of three different aspects of foundational learning: access, quality and improving data. They were asked to rank their answers between not important (score=0) and very important (score=5).

Table 17 Importance of improving access, quality and data on foundational learning

Importance	Access (# of answers)	Quality (# of answers)	Improving data (# of answers)
All respondents			
0 – Not important	0	0	1
1	0	5	6
2	5	4	5
3	9	12	10
4	15	9	12
5 – Very important	32	32	26
Not applicable	2	1	3
Average score	4.21	3.95	3.73
Count	63	63	63
Government			
0 – Not important	0	0	0
1	0	0	0

2	0	2	0
3	4	2	3
4	2	2	5
5 – Very important	19	19	16
Not applicable	0	0	1
Average score	4.60	4.52	4.54
Count	25	25	25
Development partners			
0 – Not important	0	0	0
1	0	1	2
2	3	2	5
3	1	8	5
4	10	5	5
5 – Very important	10	8	7
Not applicable	0	0	0
Average score	4.13	3.71	3.42
Count	24	24	24

13. Respondents who introduced a value of 3 or lower in the question on ‘improving data’, where prompted with an open question about on the reason why they did not think ‘improving data’ was important. The team reviewed and categorised 11 individual answers along a number of themes. The table below provides a summary of the answers.

Table 18 Reasons why improving data was not considered important

Code/theme	# of answers
Competing priorities & limited resources	5
Low Political Will or Incentives	10
Weak institutional systems & governance	7
Poor data use, culture & capacity	5
Belief Data Already Adequate / Not a Priority	1

Improving the quality of foundational learning data

14. Survey respondents were presented with a range of statements on ways to improve the quality of the data on foundational learning. They were asked to rank each statement between not important (score=0) and very important (score=5). The answers are presented in the figures below for all respondents and the two main groups of respondents (government and development partners). Not all respondents ranked all the items. Each item got between 59 and 62 replies.

Table 19 Ways to improve quality of foundational learning

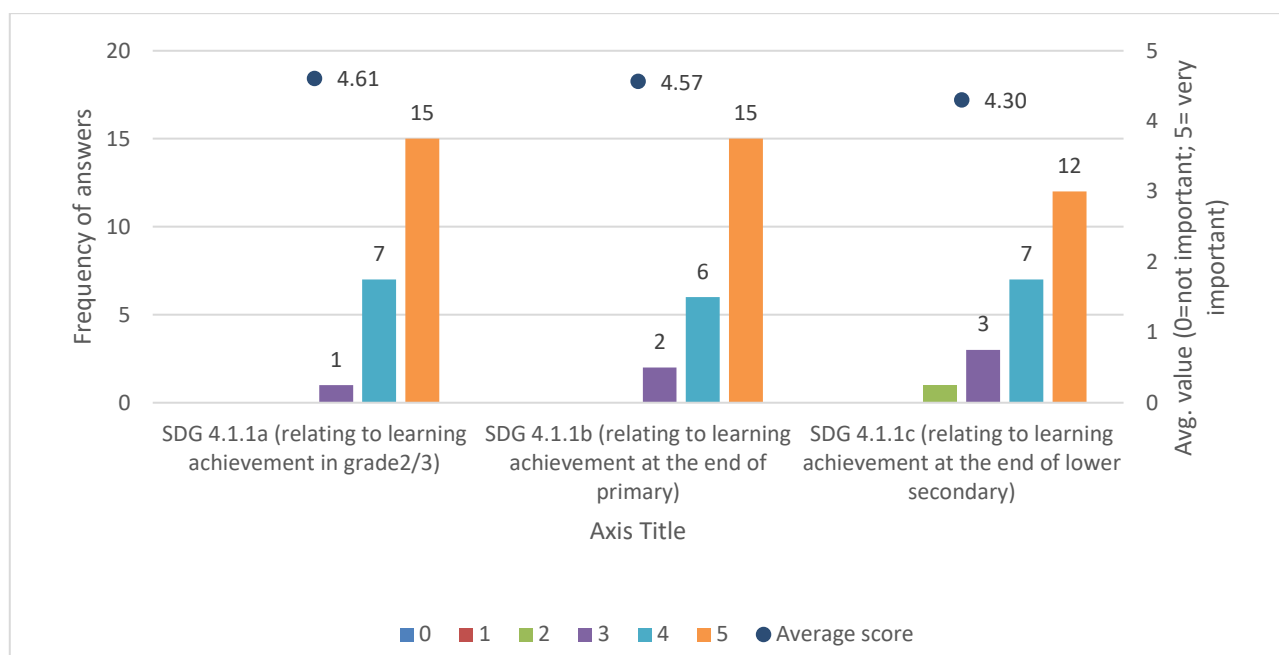
Importance	Data could be collected more frequently or regularly	More grades of pre-primary education could be assessed	More learning domains could be assessed	Data could be more representative of the population	Data could be more accurate or trustworthy	Data could be easier to compare across grades or points in time	Data could include more excluded groups children	Data could be made more available or reported more widely	Data could be easier to compare across countries	Data could be better aligned with policy needs	Education stakeholders could be helped to better use data on foundational	Standardised assessments could be complemented with assessments for
All respondents												
0 – not important	2	8	4	4	2	0	2	1	3	1	1	2
1	1	6	5	4	5	3	1	3	1	4	1	2
2	11	6	9	6	3	3	4	2	7	4	1	6
3	9	14	12	10	3	8	8	9	7	5	6	7
4	10	9	7	10	14	18	13	12	11	13	13	13
5 – very important	28	17	22	24	32	25	32	35	29	32	38	32
Not applicable	0	2	0	0	1	2	0	0	2	0	2	0
Average score	3.77	3.02	3.34	3.55	4.00	4.04	4.08	4.15	3.88	4.05	4.38	3.98
Count	61	62	59	58	60	59	60	62	60	59	62	62
Government												
0 – not important	1	1	1	1	0	0	0	0	0	1	0	0
1	0	1	0	1	1	1	0	2	0	0	0	1
2	4	2	3	2	2	2	1	1	4	0	1	2
3	4	4	4	5	2	4	2	4	3	4	2	2
4	2	4	5	4	4	5	6	3	3	3	6	4
5 – very important	13	12	12	11	15	11	15	15	13	14	16	16
Not applicable	0	0	0	0	0	0	0	0	1	0	0	0
Average score	3.88	3.88	3.92	3.79	4.25	4.00	4.46	4.12	4.09	4.27	4.48	4.28
Count	24	24	25	24	24	23	24	25	24	22	25	25
Development partners												
0 – not important	1	5	2	2	1	0	1	1	3	0	1	2
1	1	5	5	3	4	1	1	1	1	3	1	1
2	5	2	4	3	0	1	3	1	2	3	0	3
3	2	6	6	3	1	4	5	4	4	1	3	3
4	5	2	2	4	5	9	3	5	4	5	6	6
5 – very important	9	3	3	7	11	7	10	12	8	11	13	9
Not applicable	0	1	0	0	1	1	0	0	1	0	0	0
Average score	3.57	2.17	2.45	3.14	3.73	3.91	3.65	3.96	3.32	3.78	4.13	3.54
Count	23	24	22	22	23	23	23	24	23	23	24	24

Importance of reporting on SDG 4.1.1

15. Respondents working for government were asked about how important it is for their government to report on the different indicators under SDG 4.1.1. They were asked to rank each

statement between not important (score=0) and very important (score=5). A total of 23 replies were submitted.

Figure 17 Perceptions on the importance of reporting on SDG 4.1.1



16. Respondents were also asked about the main perceived barriers in meeting the technical criteria for reporting on SDG 4.1.1. The replies were categorised using themes. The answers are summarised in the table below.

Table 20 Barriers to reporting

Barrier	# of replies
I don't know	3
Funding	10
There are no barriers/Country already reporting or planning to report	4
Technical capacity	5
Stakeholder support for reporting	1

Country specific questions

17. Responders from countries where D4FL partners are implementing activities/initiatives supported by D4FL were prompted with a range of questions:

- Awareness of the activity/initiative supported by D4FL.
- Expected benefits.
- Negative consequences/if any.
- Overlaps with other activities/initiatives.

18. A total of 49 replies were submitting, covering 15 out of the 30 countries where D4FL partners are implementing activities. The average number of replies per country is 3.3.

19. The analysis was not conducted at country level, instead, the answers for different partners and programmes were aggregated. The activities initiatives covered were: ACER UK, PAL Network, WB's ALMA and WB's Accelerator 2.0. The answers are summarised in the table below.

Table 21 Stakeholder views on ACER UK

Question	Yes	No	Total	% Yes
Awareness				
Are you aware of the work that ACER is doing...	13	2	15	87%
Benefits				
No benefits expected	0		13	0%
Other	1		13	8%
More inclusive data on foundational learning (i.e. including previously excluded groups of children)	5		13	38%
More use of data to inform practice (e.g. in the classroom, in teacher training)	7		13	54%
Higher quality data on foundational learning (e.g. more representative, more accurate, more relevant)	8		13	62%
More attention to foundational learning	9		13	69%
More data on foundational learning (e.g. more regular data, data for more grades, data for more learning domains)	9		13	69%
More use of data in policy-making	11		13	85%
Data that is more relevant for in-country decision-making	11		13	85%
Data that is internationally comparable	12		13	92%
Negative consequences				
Do you think that the work of ACER will have any negative consequences?	0	13	13	0%
Negative consequences				
In your opinion, are there any overlaps between the work of ACER and other initiatives in your country?	2	11	13	15%
Can you indicate the name of the other initiative/s and the nature of the overlap?	Malawi - "WB work on assessments. This has now been defined to build on the ACER work." Rwanda - "UNESCO work on AMPL - not really an overlap but both are requiring time of a very busy agency"			
Do you have any other feedback on how the work of ACER has been delivered so far?	All very positive feedback - on quality, professionalism, responsiveness, approach			

Table 22 Stakeholder views on PAL Network

Question	Yes	No	Total	% Yes
Awareness				
Are you aware of the work that the PAL network is doing...	12	8	20	60%
Benefits				
No benefits expected	1		12	8%
Other	2		12	17%

Data that is internationally comparable	2		12	17%
More use of data to inform practice (e.g. in the classroom, in teacher training)	2		12	17%
Higher quality data on foundational learning (e.g. more representative, more accurate, more relevant)	3		12	25%
More attention to foundational learning	4		12	33%
More use of data in policy-making	4		12	33%
Data that is more relevant for in-country decision-making	5		12	42%
More data on foundational learning (e.g. more regular data, data for more grades, data for more learning domains)	7		12	58%
More inclusive data on foundational learning (i.e. including previously excluded groups of children)	7		12	58%
Negative consequences				
Do you think that the work of PAL Network will have any negative consequences?	4	8	12	33%
Please indicate the negative consequences	<p>Implementing partner can be a polarising force/opposed by government (n=1 - Pakistan); Undermine government data (n=1 - Nepal)</p> <p>Other two are not actually negative consequences - assessment needs to be aligned with GPF (n=1 - Pakistan) and hard to make work relevant to government (n=1 - Nepal)</p>			
Negative consequences				
In your opinion, are there any overlaps between the work of PAL Network and other initiatives in your country?	5	7	12	42%
Can you indicate the name of the other initiative/s and the nature of the overlap?	<p>Risk of overlap with Government initiatives (n=3) [Nepal - government-funded EGRA; Nepal - "all existing national systems" - better to strengthen those; Tanzania - NECTA 3Rs initiative]; Other funded research initiatives (n=2) [Pakistan - nationally representative achievement test funded by FCDO which will incorporate FNL and expected to report internationally; Kenya - large-scale exercises funded by Gates, TARL implementers]; Other CSOs (n=1) [Tanzania - other CSOs assessing learning more holistically]</p>			
Do you have any other feedback on how the work of PAL Network has been delivered so far?	<p>Transcend individual projects (n=2) [Mexico - transcend specific projects, integrate into public policies; Nepal - approach is "projectized", move towards institutionalisation with government]; Lack of visibility (n=2) [Nepal - work is not visible; activities not promptly communicated to government]; Valuable in highlighting importance of measuring FNL (n=1) [Mexico - highlighted importance, promoted use of data]; Valuable source of data (n=1) [Pakistan - year-on-year, in and OOSC]; Strengthen relationship with governments (n=1) [Mexico - to move towards institutionalisation; Potential to strengthen tools (n=1) [Kenya - use psychometrics]; have a designated point of contact (n=1) [Senegal - to facilitate comms and work planning]</p>			

Table 23 Stakeholder views on WB-ALMA

Question	Yes	No	Total	% yes
Awareness				
Are you aware of the work of by the World Bank funded through Accelerating Learning Measurement for Action (ALMA)...	16	2	18	89%
Benefits				
No benefits expected	0		16	0%
Other	2		16	13%
More inclusive data on foundational learning (i.e. including previously excluded groups of children)	9		16	56%
More use of data to inform practice (e.g. in the classroom, in teacher training)	9		16	56%
Higher quality data on foundational learning (e.g. more representative, more accurate, more relevant)	14		16	88%
More attention to foundational learning	11		16	69%
More data on foundational learning (e.g. more regular data, data for more grades, data for more learning domains)	13		16	81%
More use of data in policy-making	14		16	88%
Data that is more relevant for in-country decision-making	14		16	88%
Data that is internationally comparable	13		16	81%
Negative consequences				
Do you think that the work of WB/ALMA will have any negative consequences?	0	16	16	0%
Negative consequences				
In your opinion, are there any overlaps between the work of WB/ALMA and other initiatives in your country?	2	14	16	13%
Can you indicate the name of the other initiative/s and the nature of the overlap?	Both positive - WB Boost "well-aligned" (Tanzania) and ALMA seen as complementary to ACER work (Malawi)			
Do you have any other feedback on how the work of WB/ALMA has been delivered so far?	Positive feedback - WB a valued partner, work on capacity building and FNL valued. ¹¹¹			

Table 24 Stakeholder views on WB-Accelerator 2.0

WB Accelerator	Yes	No	Total	% yes
Awareness				
Are you aware of the work of by the World Bank funded through the Accelerator programme...	18	9	27	67%
Benefits				
Strengthen government systems to deliver effective teacher professional development programs	0		18	0%

¹¹¹ A respondent comment to this question was redacted to avoid readers misinterpreting the comment as a FCDO opinion or an evaluation team opinion.

More attention to foundational learning	0		18	0%
More data on foundational learning (e.g. more regular data, data for more grades, data for more learning domains)	0		18	0%
Higher quality data on foundational learning (e.g. more representative, more accurate, more relevant)	0		18	0%
More inclusive data on foundational learning (i.e. including previously excluded groups of children)	0		18	0%
Data that is more relevant for in-country decision-making	0		18	0%
Data that is internationally comparable	0		18	0%
More use of data in policy-making	0		18	0%
More use of data to inform practice (e.g. in the classroom, in teacher training)	0		18	0%
No benefits expected	0		18	0%
Other	3		18	17%
The use of appropriate language of instruction	6		18	33%
Strengthening system support for instructional coherence	7		18	39%
Improved curriculum and curricular goals for foundational learning	10		18	56%
Improved learning assessments	12		18	67%
Improved textbooks and teaching and learning materials	14		18	78%
Improved teaching in the classroom (including teacher training and support to teachers)	15		18	83%
Negative consequences				
Do you think that the work of WB/Accelerator will have any negative consequences?	0	17	17	0%
Negative consequences				
In your opinion, are there any overlaps between the work of WB/Accelerator and other initiatives in your country?	5	12	17	29%
Can you indicate the name of the other initiative/s and the nature of the overlap?	UNICEF, GPE, Prevail fund, other NGOs and WB boost programme mentioned. Most neutral but one comment that the remedial space is "hugely crowded and poorly coordinated". On the other hand, comment about Prevail fund notes process to ensure they are complementary.			
Do you have any other feedback on how the work of WB/Accelerator has been delivered so far?	No clear themes however some relate to coherence and duplication - "What appeared to be an overlap was resolved when educational authorities invited the various partners to align themselves with the programme through integrated actions and a common name" (Cote de Ivoire), could be better joined up with other FCDO			

	programmes specifically WWHGE (Sierra Leone, Rwanda),
--	---

Annex 10 Programme performance against the assessment criteria

1. This annex assesses each of the criteria in the evaluation framework. We provide two ratings for each criterion:
 - Mid-term judgement: Criteria are RAG-rated:
 - Green where they have been achieved or are very likely to be achieved
 - Amber where they are at risk or likely to be only partially achieved
 - Red where they have not or are unlikely to be achieved.
 - Grey where it is too early to draw conclusions on that criterion or the evaluation team has insufficient evidence to date to make a judgement.
 - Strength of evidence: Provides the evaluation team’s judgement on the generalisability of the evidence and how well triangulated the finding is:
 - A is well triangulated (evidenced by a majority of stakeholder KIIs/FGDs, and/or by credible documentary evidence, and or by credible quantitative evidence) and generalisable to the full programme (if triangulated in relation to specific cases only)
 - B is triangulated (supported by many stakeholder KIIs/FGDs, and/or by credible documentary evidence, and/or by credible quantitative evidence) and is generalisable to the full programme (if triangulated in relation to specific cases only), but with some contrary evidence (from any credible source)
 - C signifies significant gaps in triangulation (supported by very few or only one stakeholder KII/FGD and credible documentary evidence or quantitative evidence was not available), or significant contrary evidence was also found, but applies to full programme
 - And D signifies significant gaps in evidence or significant contrary evidence was also found (supported by very few or only one stakeholder KII/FGD and credible documentary evidence or quantitative evidence was not available), and applies to few or one selected cases, but the team judges it to be an important insight.
2. The sources of evidence for each evaluation criterion are set out in Annex 7.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
Relevance				
1. To what extent does D4FL address the needs of LICs/LMICs?				
1.1 To what extent are the programme impacts and objectives relevant to the needs of LICs/LMICs?	Improved foundational learning outcomes is a priority need of LICs/LMICs.	Green	A	In absolute terms a priority because of the gap, but also evidence that countries make it a priority
	Generating comparable and disaggregate learning data to report globally on learning progress is a priority need of LICs/LMICs.	Green	A	Country case study evidence and survey evidence, but to note that data for national use and use capability were consistently seen as more important
	Generating comparable learning data that can be used in decision-making and the design and monitoring of learning interventions is a priority need of LICs/LMICs.	Green	A	Evidence from country cases and survey that it is a need of governments, and a need for countries overall given that many do not have data
	Strengthening capabilities to use learning data is a priority need of LICs/LMICs.	Green	A	Survey evidence, country case study evidence
	Governments and other in-country stakeholders are interested in working with D4FL to address their learning data needs.	Green	A	Demand driven programme, ACER take-up continues to grow. WB ALMA has more demand
1.2 To what extent are programme activities relevant and	The D4FL's choice of partners is relevant to the needs and contexts of LICs/LMICs.	Green	A	ACER expertise relevant, WB funding power makes it a relevant partner. PAL Network relevant to countries, if not always seen as such by governments. NFER relevant because of expertise to support the KNEC, and UNICEF relevant where countries do not have data.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
aligned to the needs and contexts of LICs/LMICs'?	Implementing partners have appropriate processes in place to assess in-country stakeholders' needs and the context, and tailor their support, including when/if circumstances change.	Green	D (Significant gaps as no evidence yet for UNICEF and NFER, 2 of 5 partner evaluated)	ACER UK undertakes scoping exercises at start of support; WB TTLs' insights on support needed is built into their ongoing support; NFER has committed itself to respond to countries' needs, but no evidence yet; UNICEF no evidence either way, and PAL Network works closely with country partners
	D4FL-funded activities respond to in-country stakeholders' needs (including adapting to changing needs) and are aligned to context, including country policies, plans, education structures, logistical features (e.g. timing and mode) of assessment activities and capacities.	Green/Amber	B (D for UNICEF)	ACER UK and PAL Network (although not in relation to government) green, UNICEF no evidence yet, NFER green, WB amber – evidence both that TTL adjust (e.g. Somalia, Ghana) and also only to a point (Nepal) even if it means less connection to country processes.
	D4FL supports the right functions within countries' assessment systems to produce learning data to meet global and in-country needs (systems include assessment for learning, examinations, citizen-led assessments, sample-based standardised assessment).	Amber	A (D for UNICEF, but as only one country judged not to be a significant gap)	Good evidence that D4FL works closely with the right assessment units to produce globally comparable data, but also evidence that there is no provision in design of programme activities to extend support beyond these units towards the D4FL <i>Use</i> objective.
2. To what extent do D4FL programme's objectives and activities address the needs of the global community?				
No sub-	Improved foundational learning outcomes remains a global priority.	Green	A	KIIs with D4FL programme partners and other members of the GAML.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
questions	There is a continued global commitment to filling global learning data gaps to monitor education quality and learning outcomes.	Green/Amber	A	Countries think global benchmarking of learning important, as well as global reporting, but less important than other areas of improvement. Commitment still by UIS/GEM-R. Global assessment stakeholders noted that county assessment systems should be strengthened first. KIIs with global assessment stakeholders and in case study countries, survey
Coherence				
3. To what extent is D4FL a coherent programme across strands and partners?				
No sub-questions	FCDO, implementing partners and countries are working towards common objectives, in particular the D4FL programme objectives, namely filling global data gaps, the availability of more disaggregated data that can be used, and increased country capability to use learning data to design and monitor foundational learning interventions.	Green	A	D4FL partners share the aim of improving learning outcomes and the objective of increasing the availability of (internationally comparable) foundational learning data. While there is little investment in data use, partners are generating/supporting the generation of data with the aim that it be used to intervene on foundational learning.
	FCDO and implementing partners present a coherent narrative about the value and complementarity of different types of data.	Red	B	There are incoherencies around a) the relative emphasis placed on international reporting vs. in-country use; and b) the merits of data generated outside of government.
	FCDO and implementing partners take appropriate and effective steps to link up and ensure their work is complementary and not duplicating at global and country level.	Amber	B	The evaluation has not found evidence of substantial duplication. However, opportunities to join up work and ensure that is complementary and meets data use objectives are being missed.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
4. To what extent are D4FL activities coherent and complementary with other global learning data efforts?				
No sub-questions	D4FL is adding value to the international learning data landscape while not duplicating efforts.	Green	A	The evaluation found no evidence of duplication of other efforts, and there are some strong examples of the D4FL work being tailored to complement and support other initiatives.
	D4FL is not duplicating other international learning assessment initiatives at country level.	Green	B	The evaluation found evidence of complementarity, e.g. ALMA working alongside regional assessment programmes such as PASEC and SACMEQ. There is no explicit evidence of duplication in case countries, but surveys point to many assessments done in the past in D4FL countries.
	D4FL contributes to increased harmonisation and coordination across global efforts to improve the tracking, generation and use of foundational learning data.	Green	A	Amid ongoing inconsistencies within the global learning data sector, D4FL has contributed to greater harmonisation of global learning assessment approaches, supporting dialogue and helping align standards across implementing partners and global actors.
5. To what extent is D4FL coherent and complementary with other FCDO programmes?				
No sub-questions	D4FL and other centrally managed FCDO programmes (CMPs), in particular the What Works Hub for Global Education (WWHGE) and Scaling Access and Learning in Education (SCALE) are cooperating at the global level to strengthen learning outcomes in LICs/LMICs.	Green	B	D4FL and other FCDO CMPs are designed to be complementary, and expected to link up through the Learning at Scale pathway. This framework demonstrates clear potential for impact and global leads are committed to cooperating to amplify impact.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
	Activities in the same country and funded by D4FL and other centrally managed FCDO programmes, in particular the WWGE and SCALE, ¹¹² are well-coordinated and complementary towards improved learning outcomes.	Amber/Red	B	Despite the strong theoretical links between CMPs and bilateral programmes, evidence of this intended complementarity translating into practice in countries remains limited and inconsistent
	D4FL activities are supported by and support bilateral aid programmes to achieve D4FL and FCDO objectives.	Red	B	The evaluation has found no evidence of D4FL activities linking with bilateral programming to amplify impact.
Effectiveness				
6. To what extent is D4FL leading to the availability of more disaggregated learning data in LICs/MICs, that can be used to design and monitor foundational learning interventions?				
6.1 To what extent is D4FL increasing the availability, quality, and/or comparability of foundational learning data in LICs/MICs?	D4FL-funded activities have led (or are likely to lead) to more data on foundational learning being available in supported countries.	Amber	D	Evidence that more data available three of six country cases. In most other cases data are improved (or likely to be improved) in terms of coherence and quality, but not necessarily more data. However, outside of country cases insufficient evidence of extent of data already available.
	D4FL programme activities have improved (or are likely to improve) the quality of learning data available in supported countries in one or more of the following dimensions: <ul style="list-style-type: none"> • Frequency or regularity of available data • Number of grades of primary education that are assessed 	Green	A	See Box 2.

¹¹² Other relevant CMPs are FCDO support for the Ed Tech Hub, and the Education in Emergencies Programme.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
	<ul style="list-style-type: none"> • Number of learning domains that are assessed (literacy, maths, other aspects of learning) • Data completeness, i.e. Increased or better population representativeness¹¹³ • Data accuracy, i.e. the data is likely to be a reliable reflection of reality • Improvement in perceptions of data credibility and trustworthiness. 			
	<p>D4FL activities have improved (or are likely to improve) the comparability (temporally, spatially) of learning data in supported countries.</p>	<p>Amber</p>	<p>D</p>	<p>Yes, Ghana, more temporally; Somalia, more spatially (if national assessments get off the ground) in the country cases. Senegal yes, of region supported relative to other regions. Also in Rwanda and Malawi that we already know about, but this evidence not triangulated. Gaps in evidence of extent of this improvement for ALMA countries other than Somalia and Nepal.</p>

¹¹³ Note that progress on disaggregation by population groups (other than by location) is assessed under the next sub-question.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
<p>6.2 To what extent is D4FL leading to more and better learning data for children affected by drivers of exclusion from learning?</p>	<p>D4FL activities have led (or are likely to lead) to greater inclusion in learning assessments of one or more of the following often-excluded population groups in supported countries:</p> <ul style="list-style-type: none"> • Girls/Boys (depending on context) • Children with disabilities • Children affected by emergencies and protracted crises, especially refugee children and internally displaced children • Out-of-school children. 	<p>Amber/Red</p>	<p>A</p>	<p>Only UNRWA support represents more inclusive data for displaced children. Working in emergency contexts elsewhere (Somalia, Mali, Kenya) but only on formal school-based assessments, which might cover some displaced children, but does not specifically seek to include them.</p> <p>PAL Network and MICS might also include IDPs, as household-based assessments, but no focused effort in refugee camps</p> <p>PAL Network and MICS for OOSC, rural children</p> <p>No specific efforts to include children with disabilities funded by D4FL</p>
	<p>D4FL activities have improved (or are likely to improve) availability of data that can be used to compare learning gaps between one or more of these often-excluded population groups and their peers in supported countries.</p>	<p>Red</p>	<p>A</p>	<p>No activities other than UNRWA</p>

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
6.3 To what extent is D4FL leading to learning data that can be and is useful for application and decision-making in-country?	<p>D4FL activities have led (or are likely to lead) to increase in the availability of data that:</p> <ul style="list-style-type: none"> addresses issues identified by national stakeholders as strategically important addresses the specific learning needs and gaps within the setting can respond to requests from national policy makers is easily understood by decision makers and other relevant actors is available in the right format at the right time can inform decision making on education resourcing, goals or foci can be used to advocate for action on foundational learning. 	Green/Amber		<p>In 13 of the 24 countries, regions and UNRWA where D4FL support interventions, the scope of work includes one of these aspects</p> <p>PAL Network advocacy strategy to increase likelihood that data are used</p> <p>UNICEF Advocacy component to use data for advocacy on interventions.</p>
7. To what extent is D4FL leading to better capability in priority countries to use learning data to design and monitor foundational learning interventions?				
7.1 To what extent is D4FL leading to increased capability in supported countries to use	<p>D4FL activities have built or (are likely to build) the capability of education authorities in supported countries to effectively use learning data through one or more of the following pathways,¹¹⁴</p> <ul style="list-style-type: none"> Building sufficient ownership of learning data 	Green/Amber	A	<p>Where D4FL supports countries national assessments in or other way, there is evidence of ownership, and of D4FL partners working with assessment units to build their capacity for analysis. Still little evidence D4FL is building appetite of authorities to use data and its analysis.</p>

¹¹⁴ Criteria for this question would be influenced by D4FL response to the recommendation in the formative evaluation, on how to incorporate D4FL ambitions on assessments in classroom in the ToC.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
learning data in the design and monitoring of foundational learning interventions?	<ul style="list-style-type: none"> • Building sufficient knowledge of the assessment design, and sufficient knowledge of the assessment’s strengths and limitations • Building sufficient in-country capacity to analyse assessment data to answer decision-makers’ questions • Building the appetite of education authorities to use learning data and its analysis 			
	D4FL programme activities have built (or are likely to build) the capability of non-state actors in supported countries to effectively use learning data.	Amber	A	In only two of the country support cases, other than PAL countries where this support is built into the activities, is there evidence that D4FL works beyond assessment units to build capacity for use.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
7.2 To what extent do policies, targets and interventions in supported countries use or intend to use data generated through D4FL programme activities to strengthen action on foundational learning?	<p>D4FL programme activities have led (or are likely to lead) to country authorities at different levels acting on learning, including,</p> <ul style="list-style-type: none"> Using learning data to design, implement and monitor foundational learning interventions (such as strengthened curricula, teacher development, pedagogy and teaching materials) Increased budgets for foundational learning improvement, and foundational learning, in national budgets Increased focus on foundational learning in country, in policies and other documents. 	Amber	D	<p>Some evidence of D4FL programme activities likely to lead to country authorities using learning data to design, implement or monitor learning interventions (e.g. Rwanda, Mali)</p> <p>Some evidence of D4FL programme activities being intended to support resource allocation in education (e.g. Malawi), or action plans (e.g. São Tomé and Príncipe), etc But this is only in a third of the countries supported. And, evidence of the extent to which this will happen incomplete</p>
8. To what extent and how is D4FL filling global learning data gaps?				
8.1 To what extent is D4FL programme leading to an increase in the number of supported countries reporting learning data globally?	<p>D4FL-funded actions have led (or are likely to lead) to supported countries reporting, or reporting more comprehensively, against SDG 4.1.1a or SDG 4.1.1b.</p>	Amber		<p>D4FL-supported data likely to be produced for 27 countries, but not yet clear whether PAL network data can be used (even if meeting the criteria) and whether other countries' data will pass vetting process. Of the 27 countries, 19 have reported previously to UIS, but not under the UIS criteria. Amber rating is because of uncertainty about outcome of vetting processes.</p>
	<p>D4FL-funded activities have led (or are likely to lead) to supported countries reporting data that is recognised as internationally comparable by key assessment and education actors.</p>	Green	A	<p>Same as above: but likely that all if not most of the data will be recognised by actors (e.g. UNICEF in its proposed database) even if not by UIS.</p>

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
<p>8.2 To what extent are partners working with countries to agree support packages in a timely way?</p>	<p>D4FL is supporting the development of new methodologies to generate comparable, disaggregated data.</p>	<p>Green</p>	<p>A</p>	<p>D4FL has supported the integration of the UIS criteria in MICS and ICAN/ICAR, as well as tools for integrating them into national assessments via ACER UK and the World Bank ALMA. It is not clear whether the ALMA tools when integrating into national assessments will be made available. ACER UK has said it will make its tools etc available online, as soon as first countries have passed vetting.</p>
	<p>D4FL-funded actions are demonstrating cost-effective and replicable ways of generating comparable and usable learning data, that can be taken up or scaled in practice in most LICs/LMICs (including countries affected by emergencies and fragile states), given assessment expertise and resources.</p>	<p>Green/Amber</p>	<p>A</p>	<p>The AMPL/MICS route is cost-effective and replicable, but has less benefit for supporting national assessment systems. The ACER UK/ALMA global TA route is less replicable, and can be costly, it also draws on scarce TA capacity. The NFER support for a regional hub of expertise is a third route to make capacity available to countries, and support peer learning.</p>
	<p>D4FL-supported global public goods (or other outputs) on the generation, analysis, presentation and, use of learning data for education decision-making and the design and monitoring of foundational learning interventions:</p> <ul style="list-style-type: none"> • are/will be widely disseminated by the FCDO, D4FL partners and other actors • are/can be used by non-D4FL supported global education actors and/or non-D4FL LIC/LMIC country authorities 	<p>Green</p>	<p>A (D for UNICEF)</p>	<p>ACER UK has committed to making its tools and guidance available online; the PAL Network will make its databases and the ICAN/ICAR assessment available. World Bank ALMA has not committed to making its tools to integrate in to national assessments available. Re UNICEF, we are unclear to what extent MICS methodology will be online for use by other actors. World Bank Accelerator is providing public goods for use.</p>

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
	D4FL-supported models of TA are/can be used by other assessment actors to build the capacity of education authorities in LICs/LMICs to generate and use learning data for decision making and the design and monitoring of foundational learning interventions.	Green /Amber	A	Overlapping with second criterion under this question: all the methodologies require TA.
Efficiency				
9. How well is D4FL allocating resources to maximise the chance that outcomes were achieved?				
9.1 To what extent has the FCDO allocated resources or is likely to allocate remaining resources to maximise the chance that outcomes are achieved?	The D4FL programme has sufficient resources (monetary and non-monetary) relative to its scope and methods to make a difference.	Green	A	While not directly assessed, the combination of effectiveness with the efficiency questions support this statement.
	FCDO has selected the right combination of implementing partners and allocated monetary resources between them optimally to maximise the chance that outcomes will be achieved.	Green/amber	B	While there is evidence that this is generally the case, the report discusses some tensions and data gaps in relation to the allocation between global reporting and country-level work. Similarly, there are some gaps in the understanding of how data is being used to support decision-making.
	D4FL resources leverage other resources of partners.	Green	A	This builds on finding about leveraging as well as the investment on scalable and replicable solutions
	D4FL contracts include appropriate processes to negotiate/influence and monitor how D4FL implementing partners use resources to maximise the chance that outcomes are achieved.	Green/amber	B	This is the case across most partners, but influence on the WB is weaker. There is scope to improve consistency and utility of financial reporting.
9.2 To what extent is D4FL implementing partners allocating	Implementing partners' planned use and use of funding is only for activities that can be expected to contribute to D4FL targeted outcomes.	Green/amber	A	Funding is mostly being used for activities supporting D4FL objective. Accelerator 2.0 contributes to the overall impact, but its design does not support directly any of the D4FL outcome objectives (<i>Track, Generate or Use</i>).

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
resources to maximise the chance that outcomes are achieved?	Implementing partners have allocated resources and have processes in place to plan activities in the right combination and sequence to maximise the chances that D4FL outcomes are achieved.	Green	B	This criterion overlaps with the over above and with those under efficient delivery below
10. To what extent are partners delivering their activities in an efficient way?				
10.1 To what extent are D4F implementing partners using resources in a <u>timely</u> way?	Implementing partners have delivered activities to date in a timely way.	Amber	A	Delays have been reported by multiple partners, often due to factors beyond partners' control
	FCDO and implementing partner processes are likely to facilitate the timely use of resources and delivery of activities.	Green	B	When delays have occurred, partners have adjusted planning and project delivery. Delays are not expected to have a meaningful impact on implementation.
10.2 To what extent are D4FL implementing partners using resources <u>efficiently</u> ?	Implementing partners have used resources efficiently to date.	Green	A	There is evidence that budgets are being prudentially and judiciously managed. There is also evidence of adaptative management practices
	Implementing partners have and are prioritising the efficient use of resources.		B	
	D4FL and implementing partners have processes or protocols in place to ensure that resources are used efficiently.		B	
Impact				
11. What evidence is there that D4FL's results will lead to improved quality of education and reduced learning poverty?				
No sub-questions	Data generated through the D4FL programme is used to apply pressure at global and country level for action on foundational learning.	Amber	D	Incremental achievements yes, as in the intention to use the data to apply pressure is there, but limited evidence yet on how data will be used.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
	Commitments by global actors and LIC/LMIC education authorities to improve foundational learning can be credibly linked to the D4FL programme activities and results.	Amber	D	Limited evidence available.
	Commitments by global actors and LIC/LMIC education authorities to include or better serve populations often excluded from foundational learning can be credibly linked to the D4FL programme activities and results.	Amber/Red	A	There is limited evidence of commitments to inclusive assessments that could be attributed to these impacts of the D4FL programme.
	Actions in supported countries observed against indicator 7.2 can be expected to improve the quality of education and therefore foundational learning outcomes.	Amber	D	Evidence is incomplete, but in some cases the intent is for improvement from the D4FL actions considered in 7.2, but these actions are only in a third of countries.
Sustainability				
12. Are in-country benefits likely to be sustained?				
12.1 Are in-country changes in structures, policies and practice likely to be sustained?	Governments/national stakeholders in supported countries have ownership of D4FL-supported activities.	Green/Amber	C	There are encouraging signs that many of the structural and policy changes supported through D4FL are on a path to sustainability, particularly in countries where assessment systems were already emerging or where government ownership has been strong. See also Question 7.1
	Country capacities for assessment built through D4FL activities are sufficient and are likely to be sustained to ensure regular future assessments.	Amber	C	While there is positive evidence of capacities for assessment being strengthened in priority countries, sustainability is threatened by individuals rather than institutions being strengthened, uncertain funding, and gaps in what is left behind (software, documentation, item banks, analysis code, standard operating procedures).

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
	Country assessment frameworks, policies and targets that institutionalise learning assessment, are not likely to be overturned.	Unknown	-	Too early to assess.
	Government and/or civil society in countries supported by D4FL have committed to make financial, human, technical resources available or seek resources to repeat assessments regularly.	Amber	B	The PAL Network are committed to and fund raising for future rounds of ICAN/ICAR in the same and new countries. The evaluation has found no evidence to date of supported governments committing to fund future data generation/reporting rounds in full out of their own budgets.
	D4FL-supported country authorities who have been supported to report against SDG 4.1.1a and 4.1.1b are committed to doing so regularly.	Unknown	-	Too early to assess – no countries have yet been supported to report against SDG 4.1.1.
12.2 To what extent has D4FL generated more demand in supported countries for foundational learning data?	D4FL-funded foundational learning interventions have led (or are likely to lead) to demand from education authorities in supported countries for learning data to design and monitor these and other learning interventions, or for decision making.	Amber/Unknown	D	There is evidence that D4FL may have contributed to increased demand for foundational learning data in some supported countries, though in many cases the demand for data preceded D4FL and/or appears to be driven more by policy, donor incentives or external accountability requirements than by the programme alone. There are threats to this demand, and it remains to be seen whether it will be sustained/whether the programme will generate more demand.
	D4FL-funded interventions have led (or are likely to lead) to demand by non-state actors in supported countries for learning data to monitor foundational learning interventions, and for action on learning improvement.	Unknown	-	Too early to assess.
13. What long-term changes to the global assessment landscape are likely to result from D4FL?				

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
13.1 To what extent has D4FL contributed to clear, consistent and useful requirements for international reporting?	D4FL activities have contributed to the UIS criteria being stable so that methodologies and country capabilities developed under D4FL to report data globally, remain relevant for this purpose.	Green/Unknown	B	D4FL has played a convening role in important discussions that are believed to have led to international SDG reporting requirements that are expected to be stable and that are clear to global stakeholders. The criteria are newly released, and it is too early to assess whether they do remain stable and relevant.
	D4FL activities have contributed to the UIS criteria being clear and useful to national governments.	Amber/Unknown	C	See above re: D4FL contributions. Partners believe that the criteria will be useful to raise the quality of assessments to enable global reporting. However, there are early indications that even engaged governments will need a lot of support to achieve the criteria. Moreover, D4FL implementing partners expressed a lack of clarity about how the criteria will be applied and decisions about whether data will be accepted for SDG 4.1.1 reporting will be made.
	D4FL activities have contributed to the development of criteria for global comparability that are achievable by most LICs/LMICs.	Red/Unknown	B	See above re: D4FL’s contributions and remaining uncertainties. The level of stringency applied in the vetting process will have implications for how feasible and achievable the requirements are for LMIC/MICs, and is as yet unknown. Affordability and access to TA are also central concerns for long-term feasibility. Stakeholders pointed to the risk that compliance becomes dependent on external TA, potentially creating a model that is technically sound but not affordable. Ultimately, until the criteria are tested at scale across a wider range of contexts, their achievability and utility remain unproven.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
13.2 To what extent has D4FL contributed to methodologies that are recognised by global actors as valid to track learning in LICs/LMICs?	Global learning assessment actors (other than UNESCO) and learning experts believe D4FL-supported methodologies should be recognised as valid to track learning in LICs/LMICs.	Green	B	UIS, GPE and key donors have engaged with D4FL-supported methodologies and are beginning to integrate them into guidance, advocacy, and shared learning platforms. Stakeholders also noted that D4FL has helped elevate the legitimacy of “good-enough, credible” approaches for LICs/LMICs – an important shift in the global assessment community.
	Data generated by new D4FL-supported methodologies are used by reputable global institutions / researchers to analyse tracking progress on learning in LICs/LMICs.	Unknown	-	Too early to assess – data not generated at time of MTE data collection.
	New methodologies developed with support of D4FL have been (or are likely to be) recognised by UNESCO as valid to generate or report globally comparable learning data.	Unknown	-	Too early to assess – vetting process not in place.
13.3 To what extent has D4FL increased demand for foundational learning data outside of supported countries?	More LICs/LMICs approach UIS, FCDO and other implementing partners and assessment stakeholders for support with learning assessments and/or international reporting after seeing the outputs of D4FL.	Green	D	There are some examples of D4FL-supported approaches attracting interest from other governments and regional bodies. However, most examples relate to light-touch interest, peer learning or requests for information, rather than concrete commitments by non-D4FL countries to begin generating or reporting foundational learning data. In some cases, the demand appears to be driven more by regional or donor pressures than by internal motivation.
	D4FL partners are receiving requests for support with learning assessments and/or international reporting linked to their D4FL work.	Green	D	As above.

Question and sub-question	Judgement criteria	Mid-term judgement: RAG rating	Strength of evidence	Notes
	State and non-state stakeholders in non-D4FL LICs/LMICs want to generate and see it is as possible to generate regular foundational learning data.	Unknown	-	Too early to assess.

Annex 11 Bibliography

Short Reference	Full Reference
ACER UK, 2025	ACER UK (2025) Quarterly Report July 2025.
Addey et al., 2017	Addey, Sellar, Steiner-Khamsi, Lingard, & Verger (2017) The rise of international large-scale assessments and rationales for participation
Addey, 2024	Addey, C. (2024). Intensions, tensions and resistance: International organizations' struggles in monitoring and measuring SDG 4. <i>International Journal of Educational Development</i> , 104, 102969-. https://doi.org/10.1016/j.ijedudev.2023.102969
Creswell, 2017	Creswell, J. C. (2017). Dissemination and Reporting. In P. Lietz, J. C. Creswell, 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 and Vinjevd, 2006	Crouch, L., & Vinjevd, P. (2006) South Africa: Access before quality, and what should we do now?
Education Endowment Foundation, 2022	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 et al., 2021	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	FCDO (2022) Business Case: Data for Foundational Learning Programme
FCDO, 2024	FCDO (2024) Data for Foundational Learning Programme: 2024 Annual Report
GAML, 2018	GAML (2018) Global Alliance to Monitor Learning: 2018 Progress report
GEEAP, 2023	GEEAP (2023) Cost-Effective Approaches to Improve Global Learning: What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries? Recommendations of the Global Education Evidence Advisory Panel (GEEAP)
Görgens and Kusek, 2010	Görgens, M. G., & Kusek, J. Z. (2010) <i>Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit</i> . World Bank Publications.
Grey and Morris, 2018	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, 2022	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, 2022	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 et al., 2020	Johnstone, C. J., Schuelka, M. J., & Swadek, G. (2020) <i>Quality Education for All? The Promises and Limitations of the SDG Framework for Inclusive Education and Students with Disabilities</i> . Brill. https://doi.org/10.1163/9789004430365_004
Jóźwiak, G, 2025	Jóźwiak, G, 2025, As education funding crumbles, the sector must get its house

	in order. Devex News Education, https://www.devex.com/news/as-education-funding-crumbles-the-sector-must-get-its-house-in-order-110213
Kaffenberger, 2023	Kaffenberger, M., & Spivack, M. (2023) System coherence for learning: Applications of the RISE education systems framework. https://www.elgaronline.com/edcollchap-oa/book/9781802205930/book-part-9781802205930-18.xml
Lockheed and Wagenmaker, 2013	Lockheed & Wagemaker (2013) International Large-Scale Assessments: Thermometers, Whips or Useful Policy Tools?
Minz and Chhabra, 2023	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 and Oxford MeasurEd, 2024	Mokoro Limited and Oxford MeasurEd, 2024, Final Incept Report; Evaluation of the FCDO Data for Foundational Learning Programme, unpublished.
Montoya, 2024	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 et al., 2021	Outhred, R., Turner, F., & Mba, E. (2021) Evaluation of SEA-PLM. Cambridge: Cambridge Education.
Outhred et al., 2024	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.
R4D Institute, 2015	R4D Institute (2015) Bringing Learning to Light: The Role of Citizen-led Assessments in Shifting the Education Agenda
Ritchie et al., 2013	Ritchie, J. et al. (2013) Qualitative Research Practice: A Guide for Social Science Students and Researchers.
Rustad, 2024	Rustad, S.A. (2024) Conflict Trends: A Global Overview, 1946–2023. PRIO Paper. Oslo: PRIO.
Sayed and Moriarty, 2020	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 and Taylor, 2013	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 et al., 2016	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
Trevino et al., 2016	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
UIS, 2025a	UIS. (2025). DatabrowserUIS. Retrieved from UIS: https://databrowser.uis.unesco.org/?utm
UIS, 2025b	UIS (2025) PAL Network's ICAN/ICARe: Alignment with Global Reporting Criteria for SDG 4.1.1(a). GAML report WG/GAML/11/5.1.
UIS, 2025c	UIS (2025) Criteria for use of an assessment to report on SDG 4.1.1. Montreal: UNESCO Institute for Statistics.
UIS and ACER UK, 2025	UIS and ACER UK (2025) SDG 4.1.1 MPL. Definition and blueprint for assessment
UN, 2025	United Nations. (2025). The Sustainable Development Goals Report. United

	Nations.
UNESCO and GEMR, 2024	UNESCO and GEMR (2024) Learning. Available at: Learning GEM Report SCOPE
UNESCO, 2015	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	UNESCO (2017) Unpacking Sustainable Development Goal 4: Education 2030; guide—UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000246300
UNESCO, 2018	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	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	UNESCO (2023) 2023 SDG 4 scorecard report on progress towards national SDG 4 benchmarks
UNESCO, 2024	UNESCO (2024). Highlights from the 2024 Global Education Data Release. Retrieved from UNESCO: https://www.unesco.org/en/articles/highlights-2024-global-education-data-release?utm_source=chatgpt.com
UNESCO, 2025	UNESCO. (2025). UIS launches September 2025 Global Education Data Release. Retrieved from UNESCO: https://www.unesco.org/SDG4education2030/en/articles/uis-launches-september-2025-global-education-data-release?utm_source=chatgpt.com
UNGEI, 2023	UNGEI (2023) G7 Global Objectives Report 2023
UNICEF, 2021	UNICEF (2021). COVID-19 and School Closures: One year of education disruption.
UNICEF, 2024	UNICEF (2024) Foundational Learning Action: Findings from the Foundational Learning Action Tracker 2024
UNICEF, 2025	UNICEF (2025) Foundational Learning Action Tracker Global Report. Available at: https://www.unicef.org/media/175231/file/Foundational-learning-action-tracker-2025-Global-report.pdf
United Nations Economic and Social Council, 2023	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)
UNOCHA, 2024	UNOCHA (2024) UNOCHA Financial Tracking Service Statistics 2024. Available at: https://fts.unocha.org/countries/overview
Wagner et al., 2018	Wagner, Wolf, & Boruch (2018) Learning at the bottom of the pyramid: science, measurement, and policy in low-income countries
World Bank FLC, unpublished	World Bank FLC (unpublished.) How FLC Funded Programs Work Together: ALMA, Accelerator 2.0, and Implementation Science.
World Bank, 2024	World Bank (2024) ALMA Grant Application Form FY2025
World Bank, 2025a	World Bank (2025) Accelerator Progress Report, Oct 2025
World Bank, 2025b	World Bank (2025) Accelerator Implementation Update Report to D4FL, September 2025.

Annex 12 Acronyms

ACER	Australian Council for Educational Research
ALMA	Accelerating Learning Measurement for Action
AMPL	Assessments for Minimum Proficiency Levels
BESTA	Better Education Statistics and global Action to improve learning
BMGF	Bill & Melinda Gates Foundation
CMP	Centrally Managed Programme
D4FL	Data for Foundational Learning
EARC	Education Assessment Resource Centre
ECW	Education Cannot Wait
EdGE	Education, Gender and Equality Directorate
EDS Commission	Education Data and Statistics Commission
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EIEPC	Education in Emergencies and Protracted Crises
EQ	Evaluation Question
ER	Evaluation Report
ERICC	Education Research in Conflict and Protracted Crisis
ESP	Education Sector Plan
ET	Evaluation Team
FCDO	Foreign, Commonwealth & Development Office
FGD	Focus Group Discussion
FLAT	Foundational Learning Action Tracker
FLC	Foundational Learning Compact
FLEX	Foundational Learning Exchange
G7	Group of Seven
GALOP	Ghana Accountability for Learning Outcomes Programme
GAML	Global Alliance to Monitor Learning
GAMLA	Global Alliance for Measuring Learning Assessment
GED	Girls Education Department
GEM-R	Global Education Monitoring Report
GEOP	Ghana Education Outcomes Project
IAEG-SDGs	Inter-agency and Expert Group on SDG Indicators
ICAN/ICAR	International Common Assessment of Numeracy and Reading
ILSA	International Large-Scale Assessments
Inc	Including
INEE	International Network for Education in Emergencies
IPA	Innovations for Poverty Action
IR	Inception Report

KI and KIIs	Key informants, and key informant interviews
KNEC	Kenya National Educational Assessments Council
LARS	Learning Assessment in Rwandan Schools
Lao PDR	Lao People's Democratic Republic
LIC	Low-Income Country
LLECE	Latin American Laboratory for Assessment of the Quality of Education
LMIC	Lower Middle-Income Country
LSA	Large-Scale Assessments
MEL	Monitoring, Evaluation and Learning
MIA	Medición Independiente de Aprendizajes
MICs	Middle-Income Countries
MICS	Multiple Indicator Cluster Survey
MPL	Minimum Proficiency Level
MTE	Mid-term Evaluation
NaCCA	National Council for Curriculum and Assessment
NFER	National Foundation for Education Research
NLAF	National Learning Assessment Framework
NST	Nationally Standardised Test
OECS	Organisation of Eastern Caribbean States
OOSC	Out-Of-School Children
PAL	People's Action for Learning
PASEC	Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN
PILNA	Pacific Islands Literacy and Numeracy Assessment
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SCALE	Scaling Access and Learning in Education
SDG	Sustainable Development Goal
SEA-PLM	Southeast Asia Primary Learning Metrics
TA	Technical assistance
TaRL	Teaching at the Right Level
THRIVE	Early Childhood Development at scale
TIMSS	Trends in Mathematics and Science Study
TL	Team Leader
ToC	Theory of Change
TOR	Terms of Reference
TTL	Task Team Leader
UIS	UNESCO Institute for Statistics
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organisation

UNICEF	United Nations Children’s Fund
UNRWA	United Nations Relief and Works Agency for Palestine Refugees in the Near East
WWHGE	What Works Hub for Global Education