



United Nations
Educational, Scientific and
Cultural Organization

MAKING EVALUATION WORK FOR THE ACHIEVEMENT OF SDG 4 TARGET 5: EQUALITY AND INCLUSION IN EDUCATION

IOS Evaluation Office

IOS/EVS/PI/178



Commissioning office:

UNESCO Internal Oversight Service (IOS) Evaluation Office

Authors:

Dr. Karen Mundy

Dr. Kerrie Proulx

with research support from Dr. Caroline Manion

Reference group members:

Kathleen Letshabo, UNICEF Evaluation Office

Siv Lillestøl, NORAD Evaluation Department

Susan Cáceres, World Bank Internal Evaluation Group

Cristina Marosan Ling, World Bank Education Global Practice

Christian Stoff, Education Cannot Wait, Monitoring, Evaluation and Reporting Team

Deborah McWhinney, World Food Programme, Office of Evaluation

Astrid Gillet, UNESCO Education Sector (Executive Office)

Justine Sass, UNESCO Education Sector (Section of Education for Inclusion and Gender Equality)

Evaluation period:

January-June 2019

Location of the evaluand:

Global, with national workshops in Ghana, Guatemala, Lebanon, Nepal and Peru

Report submission:

July 2019

ABSTRACT

The adoption of the Sustainable Development Goals (SDGs) and the Education 2030 Framework for Action has placed gender parity, equality and inclusion in education at the heart of the international development agenda and is specifically formulated in SDG 4 Target 5. Now, as never before, relevant and contextualised evidence, driven in part by robust evaluation data, is needed to track and strengthen progress on educational equity. To this end, a group of international organisations, led by UNESCO, have combined efforts to explore how their education evaluations can better support Member States to strengthen progress on gender parity, equality and inclusion in education.

The following report synthesises evidence from publicly available independent evaluations from 13 organisations, highlighting evidence gaps and summarising 'what works' for whom and in what contexts for advancing gender equality and equity for vulnerable groups. The study also reports on responses to the synthesis from national stakeholders in five countries: Ghana, Guatemala, Lebanon, Nepal and Peru. The report culminates in a discussion and recommendations, which outline the need for greater coordination and collaboration in four key areas to further enhance the contribution of evaluations to global and country level progress on SDG 4 Target 5.

ACKNOWLEDGEMENTS

The Internal Oversight Service (IOS) Evaluation Office would like to acknowledge all who participated in and contributed to this study. The evaluation was managed by Mr. Moritz Bilagher, Principal Evaluation Specialist, IOS Evaluation Office, and conducted by Professor Karen Mundy from the University of Toronto and Dr. Kerrie Proulx, with research support from Dr. Caroline Manion. The support provided by Mr. Giacomo Tirelli, Assistant Programme Coordinator in the IOS Evaluation Office, was highly appreciated.

The Evaluation Office would like to thank all informants who have made this evaluation possible as well as those colleagues, inside and outside of the Organisation, who have provided feedback on the report. These include, in particular, the representatives from the partner organisations: UNICEF Evaluation Office, Norad Evaluation Department and the World Bank Internal Evaluation Group.

Susanne Frueh
Director, IOS

Table of Contents.....4

Text Boxes	5
List of Tables	5
List of Figures	5
Acronyms and Abbreviations.....	6

Executive Summary.....7

1. Introduction	10
2. Objectives, Scope and Key Questions.....	11
3. Conceptual Framework and Analytical Approach.....	12
4. Methodology.....	14
5. Evaluations by Geography, Level of Analysis, Questions and Methodologies	17
6. Types of Educational Initiatives and Target Populations in the Evaluations.....	25
7. Lessons From the Evaluations about What Works to Support SDG 4 Target 5	29
8. Country-level Responses to the Study.....	37
9. Conclusions and Recommendations.....	39
Reference list.....	42
Appendix A: List of Included Evaluations	46

Text Boxes

Box 2-1: Key questions that guide the evidence synthesis	11
Box 3-1: Dimensions of equity covered in the evaluation synthesis	13
Box 5-1: Evaluations using quantitative counterfactuals by intervention type	22
Box 5-2: Illustrative comments about data quality	24
Box 7-1: Cash transfers have a strong impact, but some limitations	29
Box 7-2: Better evidence on non-formal education is needed	30
Box 7-3: Illustrative comments on equity and access objectives in system-level interventions	33
Box 7-4: Disability is neglected	34
Box 7-5: UNICEF-UIS Out-of-School Children	34
Box 7-6: Greater coordination and peer learning exchange	35
Box 8-1: Joint evaluation as best practice in Nepal	37

List of Tables

Table 2-1: Organisations with evaluations included in the synthesis	12
---	----

List of Figures

Figure 4-1: Evaluation search and selection (PRISMA Flow Chart)	14
Figure 5-1: Outputs/outcomes reported in the evaluations	18
Figure 5-2: Level of education covered by evaluated initiatives	19
Figure 5-3: Evaluations in the dataset by organisation	20
Figure 5-4: Evaluation reports by country and region	21
Figure 6-1: Distribution of evaluations by intervention level and across intervention areas	27
Figure 6-2: Populations targeted by educational interventions	28

Acronyms and Abbreviations

3ie	International Initiative for Impact Evaluation
ADB	Asian Development Bank
DFID	Department for International Development
ECW	Education Cannot Wait
EFA	Education for All
EMIS	Education Management Information System
GEC	Girls' Education Challenge
GPE	Global Partnership for Education
HLPF	High-level Political Forum
ICAE	International Commission for Aid Effectiveness
ICR	Implementation Completion Report
ICT	Information and Communication Technology
IDB	Inter-American Development Bank
IEG	Independent Evaluation Group
IOS	Internal Oversight Service
M&E	Monitoring and Evaluation
MDGs	Millennium Development Goals
Norad	Norwegian Agency for Development Cooperation
OECD DAC	Organisation for Economic Co-operation and Development – Development Assistance Committee
PPAR	Project Performance Assessment Report
RCT	Randomised Control Trial
SDG	Sustainable Development Goal
ToR	Terms of Reference
TVET	Technical and Vocational Education and Training
UIS	UNESCO Institute for Statistics
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNFPA	United Nations Population Fund
UNGA	United Nations General Assembly
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNRWA	United Nations Relief and Works Agency
USAID	United States Agency for International Development
WFP	World Food Programme

Executive summary

1. The Sustainable Development Goals (SDGs) and the Education 2030 Framework for Action, adopted in 2015, placed equity and inclusion at the heart of the international development agenda. While targets for gender parity, equality and inclusion in education appear across SDG 4 (quality education), they are captured explicitly in SDG 4 Target 5: “By 2030, eliminate gender disparities in education for all and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations”.
2. Tackling equity challenges in education will require a collective effort to progress towards inclusive, equitable and good-quality education systems and ensure that services reach the poorest, the most vulnerable and those who are most often left behind. Now, as never before, relevant and contextualised evidence, driven by robust evaluation data, is needed to track and strengthen progress on educational equity.
3. To this end, the evaluation offices of UNESCO, UNICEF, the Norwegian Agency for Development Cooperation (Norad) and the World Bank have combined efforts to explore how evaluations in the field of education can better support Member States and strengthen progress towards SDG 4 Target 5. Leading on this effort, UNESCO Internal Oversight Service (IOS) Evaluation Office has commissioned the following study, which synthesises publicly available evaluation evidence on what works, for whom and in what contexts, to advance on gender parity, equality and inclusion of vulnerable groups. In doing so, it assesses gaps and challenges in existing evaluation evidence. The study also reports on the viewpoints of country-level stakeholders, gathered through a series of workshops in five countries: Ghana, Guatemala, Lebanon, Nepal and Peru.

Scope, approach and methodology

4. This report draws on a dataset of 147 independent evaluations that were commissioned by the evaluation units of 13 UN and other development organisations and published between January 2015 and March 2019. Evaluations were selected for inclusion when they studied an intervention that occurred in an educational setting and reported on equity objectives or results.¹
5. The synthesis was conducted in two steps. First, evaluations were coded to allow for a descriptive mapping of the evaluation dataset. Second, narrative content analysis was used to analyse evaluation evidence at four initiative levels: children, households and communities; teachers and schools; systems and sector-level; and thematic. In addition, evidence on interventions for specific disadvantaged populations was reviewed and summarised.

Findings from the descriptive mapping of the evaluation dataset

6. **What is the availability of evaluations covering SDG 4 Target 5?** Over the period between 2015 and 2019, 30-40 evaluations in the field of education related to gender parity, equality and inclusion were produced annually by the 13 organisations combined. Geographically, the largest number of evaluations was undertaken in sub-Saharan Africa. The vast majority of evaluations in the field of education identified in the search process had objectives or outcomes related to SDG 4 Target 5. The predominant focus of these evaluations was on education access or participation (rather than completion or learning) and on interventions in basic education (understood as primary and / or secondary education) and early childhood education. The measurement of equality and inclusion was often not the primary focus of evaluations when measuring the impact of interventions.

1. Agreed with UNESCO. The organisations included in the study were agreed with UNESCO Internal Oversight Service (IOS) Evaluation Office and the reference group.

7. **What questions and methodologies are used in the evaluations?** Questions were typically aimed at understanding the effectiveness of an organisation in delivering an intervention or programme of interventions. Evaluations usually answered questions related to all five OECD Development Assistance Committee (DAC) criteria, though answers to questions on efficiency (cost-effectiveness) and impact were notably weakest. The evaluations were typically comprised of desk reviews of Monitoring and Evaluation (M&E) reports and other documents, field interviews and site visits and secondary analyses of school-level, national or regional data. Only 28 evaluations used rigorous quantitative evaluation methods with a counterfactual. A challenge that emerged from this review is the need to further strengthen the evidence provided by evaluations of complex, multi-level component programmes that aim to support system-level reform. The limited availability, quality and comparability of data on equality outcomes were repeatedly noted across the evaluations as limitations. Concluding recommendations in the evaluations often focused on implementation considerations, but rarely included an explicit focus on mechanisms to improve gender equality and equity in the education initiatives under evaluation.
8. **What kinds of educational initiatives are included in the evaluations?** The largest group of interventions evaluated were directed at the level of children, households and communities (e.g., school feeding, cash transfers and non-formal life skills programmes). The second largest group of evaluations were of programmes operating at the system or sector-level. The smallest number of evaluations (and with the weakest evaluation evidence related to equality and inclusion) was found for interventions at the level of teachers and schools.
9. **What target populations are expected beneficiaries in the activities being evaluated?** A large number of the interventions evaluated targeted activities for gender (and predominantly girls) as well as activities targeting socioeconomic disadvantage and place of residence. Disability and other forms of inequality linked to ethnicity, culture or language were less frequently mentioned. An important aspect of this dataset is the large number of evaluations of interventions that address the needs of children in fragile and conflicted-affected settings, contexts that are rarely the focus of other large scale syntheses of evidence in education.

Lessons about what works to support SDG 4 Target 5

10. **The strongest evidence of impact related to equality and inclusion was found for interventions at the level of children, households and communities, especially cash transfers and school feeding programmes.** Conditional cash transfers improved school attendance among children in poor households in several contexts, as did unconditional transfers in others. They can serve to boost enrolment for girls in schools where gender parity has not been achieved and can reduce children's work and household chores. Bursaries may lead to improved access to post-secondary education for vulnerable youth and refugees. While school feeding programmes have strong positive effects on primary school enrolments, particularly in areas of high food insecurity including drought-affected and internally displaced/refugee contexts, unintended results identified include overcrowded classrooms, sometimes aggravated by the conversion of classrooms into food storage and eating spaces and additional demands on teaching staff and school management, which may detract from their core educational roles. Constraints and limitations of these interventions are also highlighted by the evaluations. Non-formal life skills programmes and community-based information and advocacy interventions show some promising anecdotal and qualitative evidence for promoting gender equality and improved opportunities for vulnerable groups (especially in conflict-affected contexts), but need a larger and more diverse pool of evidence.
11. **Evaluations at the level of teachers and schools (including pedagogy) provided limited evidence of impact related to reducing inequality,** signaling the need for stronger evaluative evidence in this area. Still, there is evidence to suggest that improved teacher training increases school completion rates, especially for girls, in some contexts. In other contexts, programmes for building new schools reduced school drop-out and improved perceptions of the quality of school, and the provision of water, sanitation and hygiene (WASH) facilities increased school enrolment in arid, pastoralist areas. While results pertaining to information and communication technology programmes (ICT), and specifically online learning, reported mixed results, interventions supporting the education of children during emergencies show considerable evidence of rapid increases in enrolments in refugee camps/host communities following relief responses.
12. **Evidence synthesised from evaluations at the system and thematic levels provided important insights.** For example, they explored the mixed impacts of new forms of development financing (e.g., results-based aid) and point to equity and inclusion challenges in complex decentralisation and school-based management

reform. The limited available evaluation evidence suggested limited impact of results-based financing on school completion, learning or gender equality. In one case, findings suggested that while budget support improved overall education spending and access, it did not improve policy dialogue in areas beyond access. Read together, evaluations at system and thematic levels suggested the need for greater coordination and learning across organisations, a finding that was echoed in the country validation workshops. There is an enormous opportunity to coordinate better across organisations to ensure both cross-agency and country-level learning and exchange. This analysis also highlighted the need for greater attention to building national capacity, including national capacity for monitoring and evaluation of education equality and inclusion.

Recommendations and conclusions

13. Since the adoption of the SDGs, significant work has been carried out to support improved data and indicators related to equity and inclusion in education; however, the large volume and range of evaluations conducted with reference to equality and inclusion in education has not been matched by a coordinated system to generate, synthesise and disseminate equality and inclusion-focused evaluation evidence. Four overarching recommendations are proposed to further strengthen and address educational equality and inclusion in evaluations by international and bilateral development organisations.
14. **Recommendation 1 - Address evidence gaps:** Agencies should work together to address areas with key gaps in evaluation evidence for SDG 4 Target 5. Such gap areas include the effects of teacher development initiatives, alternative/non-formal education, school-based management, school grants and decentralisation reforms in terms of achieving greater equality and inclusion. A particularly important imperative for future evaluations is the question of how complex system-level interventions can be better designed to reach children and youth with disabilities, as well as indigenous and ethnic minority groups. Evaluation funding should also be focused on areas where there is hopeful qualitative evidence but limited quantitative data to support scale-up of interventions.
15. **Recommendation 2 - Contribute to stronger and more consistently available data:** Good quality national data on equity and inclusion is a global good. The evaluation units of international organisations can support and contribute to national capacity by investing in the work of national partners when collecting data on equality and inclusion, and by building in funding for national actors to monitor and conduct formative evaluations. They can also join forces with national technical agencies and data units to ensure that national data systems are strengthened.
16. **Recommendation 3 - Strengthen evaluation methodologies:** Agencies should explore ways to strengthen their evaluations by i) incorporating stronger and more consistent equality and inclusion measures; ii) embedding experimental evaluations into the design of system-level programmes; iii) creating stronger standards for the evaluation of system-wide reform programmes in education, including through the use of theory-based methodologies; and iv) investing in the collection of cost-effectiveness data. Creating a common set of 'best practices' for evaluation in these areas would lead to significant advances in the level of knowledge and evidence available related to SDG 4 Target 5.
17. **Recommendation 4 - Synthesise and collaborate to make evidence more useful to national stakeholders:** Building on the momentum generated through this collaborative study, and on recent UN commitments for greater collaboration and coordination of evaluation work, development partners should foster ways to systematically coordinate, synthesise and promote learning from their evaluations. Effective strategies might include the creation of a registry of planned or commissioned evaluations in the field of education; commitment to regular international syntheses of evaluations; and routine efforts to link evaluation evidence to the wider research literature on what works. Even more importantly, evaluation offices should consider ways in which they can work more closely with governments and other national stakeholders to identify key areas with gaps in evaluation and to collaborate on all aspects of evaluations from the identification of questions and design to learning and dissemination.
18. This study provides a starting point for such future collaboration. It offers a path forward for ways in which international evaluations can more effectively support global learning among countries and their development partners, ensuring a robust evidence base to support the implementation and scale-up of effective education strategies to improve gender parity, equality and inclusion of marginalised groups in education.

1. Introduction

19. Between 2000 and 2015, strong gains were made towards achieving the global education goals set out in the Millennium Development Goals (MDGs) and Education For All (EFA) agenda. The world has largely achieved the gender parity target at the primary level of education, and has substantially increased the overall number of children attending school (UNESCO GEM, 2018). Rates of adult literacy are improving. Nevertheless, progress in advancing equity and inclusion has been uneven. The number of out-of-school children has remained broadly stagnant, with more girls than boys unable to access education. Less than half of countries have achieved gender parity in lower secondary education (UNESCO GMR, 2015). Vulnerable populations – including persons with disabilities, indigenous peoples and children in conflict-affected situations – are often unable to access basic education services. Even where access to education has improved, this has not necessarily led to improved learning outcomes for disadvantaged populations (UIS, 2018c).
20. The adoption of the Sustainable Development Goals (SDGs) and the Education 2030 Framework for Action has placed equity at the heart of the international development agenda. Concerns about equity and equality are spread through each of the SDG 4 targets, but are captured explicitly in SDG 4 Target 5: “By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations”.
21. Tackling equity challenges will require a collective effort to progress towards inclusive, equitable and good-quality education systems and ensure that services reach the poorest, the most vulnerable and those who are most often left behind. Now, as never before, relevant and contextualised evidence, driven by robust evaluation and research data, is needed to track and strengthen progress on educational equity. This is a fundamental contribution of the evaluation community but it requires organisations to provide evaluation evidence that is synthesised and accessible, and to strategically target resources where there are evidence gaps.
22. To this end, the evaluation offices of UNESCO, UNICEF, the Norwegian Agency for Development Cooperation (Norad) and the World Bank have combined efforts to explore how education evaluations can better support Member States and strengthen progress towards SDG 4 Target 5. Leading on this effort, UNESCO Internal Oversight Service (IOS) Evaluation Office has commissioned the following study, which seeks to synthesise publicly available independent evaluation evidence and explore how evaluations can better contribute to the achievement of the SDG for education and support country-level progress. The study also reports on the views of country-level stakeholders, sought through a series of workshops in five countries: Ghana, Guatemala, Lebanon, Nepal and Peru.
23. The study sets a pathway for further collaboration among the evaluation offices of international organisations. Its findings and recommendations will be shared at the UN High-level Political Forum (HLPF) to review the Sustainable Development Goals in July 2019 and in a variety of other national and international forums.

2. Objectives, scope and key questions

24. This evidence synthesis maps a large and diverse body of evaluations in order to examine the extent, range and nature of available evidence related to educational equity and inclusion. One of its central goals is to identify key evidence gaps and areas where evaluation can be used to support progress towards the achievement of SDG 4 Target 5. The report is organised around a series of guiding questions (see Box 2-1 below).

Box 2-1: Key questions that guide the evidence synthesis

Question 1: To what extent is SDG 4 Target 5 covered by the publicly available evaluations undertaken worldwide by major multilateral and bilateral aid agencies since 2015? Among those covering SDG 4 Target 5, what types of questions do the evaluations seek to answer and using what methodologies?

Question 2: According to these evaluations, what kinds of initiatives are taken to address gender disparities and other inequalities in education? What are the populations targeted by these initiatives being evaluated, e.g., gender, disability, refugee, ethnicity, rurality or socioeconomic status? Where do they fit along the theory of change from family to systems-level interventions?

Question 3: What evidence of success is found by the evaluations regarding initiatives that tackle gender disparity and exclusion? Where do the evaluations point to initiatives that have the potential for replication and what policy actions do the evaluators recommend?

Question 4: What do these evaluations tell us about key contextual factors and how they interact with these initiatives? What are the barriers and facilitators to effective implementation of these initiatives?

25. The study is comprised of a mapping and structured synthesis of evaluations and their findings. It was undertaken through a formal review of publicly available education-related evaluations commissioned or conducted by the evaluation units of 17 UN and other multilateral and bilateral organisations between January 2015 and March 2019 (see Table 2-1). It includes evaluations of initiatives conducted in educational settings which report on objectives or results (i.e. output, outcome, impact) for gender parity, equality and inclusion of vulnerable groups.²
26. The review does not include self-evaluations or validations of implementation completion reports, nor does it cover research studies conducted by these organisations. Thirteen of the original 17 organisations included in the initial search strategy were found to have evaluations meeting the inclusion criteria for this study.

² The organisations included in the study were agreed with UNESCO Internal Oversight Service (IOS) Evaluation Office and the reference group.

Table 2-1: Organisations with evaluations included in the synthesis

1. Asian Development Bank (ADB)
2. Global Partnership for Education (GPE)
3. Inter-American Development Bank (IDB)
4. Norwegian Agency for Development Cooperation (Norad)
5. UK DFID
6. UNESCO
7. UNICEF
8. UN Women
9. UN Refugee Agency (UNHCR)
10. United Nations Development Programme (UNDP)
11. UNRWA
12. World Food Programme (WFP)
13. World Bank

3. Conceptual framework and analytical approach

27. This study conceptualises equality and inclusion broadly, looking beyond equality in access to education, to school completion and learning (see Box 3-1 for definitions). This conceptualisation is in keeping with what UIS explains is the concept of equity in relation to SDG 4 Target 5, which "... broadly refers to different concepts related to fairness and compensatory actions that recognise disadvantage"(UIS, 2018a, p. 32). Therefore, included evaluations are of any interventions that reported results or had as their objective improving education outcomes for disadvantaged or vulnerable groups, including people affected by conflict, people with disabilities and people marginalised due to ethnicity, socio-economic status and place of abode (e.g., rural children).
28. **Conceptual framework:** Building on the framework employed by Snilstveit et al. (Snilstveit, Stevenson, Phillips, Vojtkova, Gallagher, Schmidt, Evers, 2015, 2016) in their comprehensive systematic review of impact evaluations of education interventions in low- and middle-income countries, evaluations were organised according to whether the intervention was focused at one of four levels:
- I. Children, households and communities
 - II. Teachers and schools (including pedagogy)
 - III. Education systems and sectors
 - IV. Thematic (global)
29. Two of these levels are different from those proposed in the Snilstveit et al. framework, reflecting the fact that many of the evaluations included in this study are of complex interventions with multiple sub-components, while others are of a portfolio of global or cross-national activities.
30. **System- and sector-level interventions:** Many of the evaluations in the dataset are of complex, multi-component activities that aim to support the reform of multiple levels of intervention of an educational system. Included at the system-level are evaluations that focus on sector-level governance, management and financing.

31. **Thematic and global-level interventions:** Evaluations of donor initiatives that are cross-national and sometimes cross-sectoral or evaluations that review an organisation's portfolio of activities on a specific theme (e.g., gender equality, peace-building through education) or within a specific level of education (e.g., early childhood, higher education).

Box 3-1: Dimensions of equality covered in the evaluation synthesis

Access: Children and youth officially registered in a given educational programme, both formal and non-formal and regardless of age, by at least one equity dimension (e.g., gender, socio-economic status (SES), refugee/displaced). The most common set of indicators in the evaluations used to measure access to an educational programme are gross or net enrolment and/or attendance rates, generally stated as ratios by equality dimension. Some evaluations included percentage of children ever accessing school and total years of schooling.

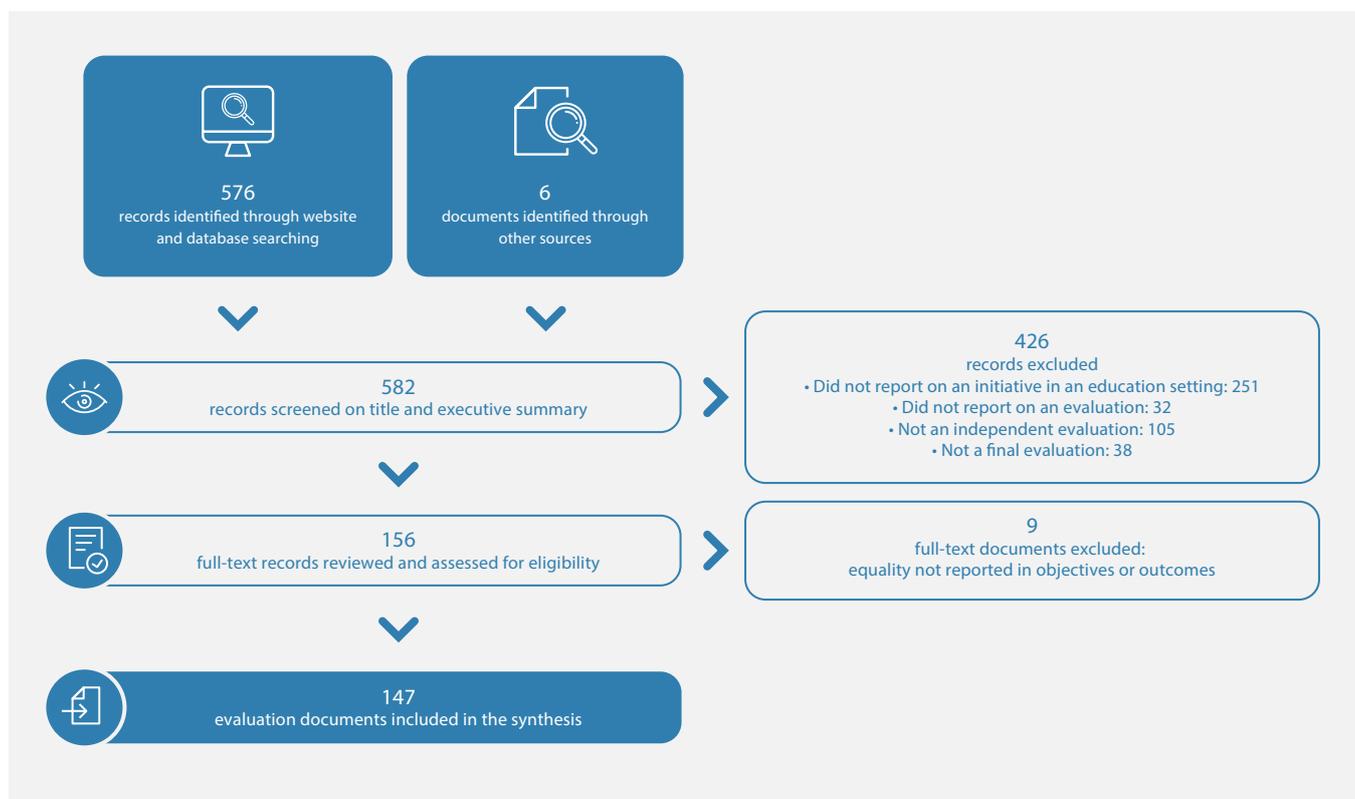
Completion: How many students have completed primary, secondary or tertiary education and progress through the education system by at least one equality dimension. Example indicators include: drop-out rate per education level or education initiative, completion rate by education level, survival rate and transition rate to next level of education.

Learning: Defined as achievement of learning objectives, both cognitive and non-cognitive. Where used, indicators consisted mainly of assessments of literacy and language skills, and mathematics, and included percentage of students passing national examinations, early learning achievement rates and psycho-social well-being.

4. Methodology

32. **Evaluation selection:** Figure 4-1 below provides an overview of the process used to select evaluations for the dataset. Only publicly available evaluations were considered in order to ensure a transparent and reproducible method. English, French and Spanish documents were included. The search process resulted in the identification of 582 potentially relevant documents, which were then carefully screened based on eligibility criteria (i.e., education setting, independent evaluation, document type, final or end-line evaluation). Self-evaluations, such as Implementation Completion Reports (ICRs), were not included, lowering the yield for some organisations (notably the World Bank). Following the screening process, the final dataset for the evidence synthesis included 147 documents. The number of included evaluations is significantly larger than originally anticipated, reflecting a substantial volume of evaluation evidence over the past few years. Appendix A provides a complete list of the evaluations included in the synthesis.

Figure 4-1: Evaluation search and selection (PRISMA Flow Chart)



33. **Data extraction:** All full-text evaluation documents were uploaded to the Evidence for Policy and Practice Information Reviewer (EPPI-Reviewer) software, designed to support systematic review of qualitative and quantitative evidence. An EPPI-Reviewer data extraction framework was used to collect and code comprehensive descriptive data on each evaluation document. To harmonise and identify elements common across the evaluations, data were coded using the SPICE Framework, consisting of five distinct features (Moher, Liberati, Tetzlaff, Altman & PRISMA Group, 2009):

- Setting (Where?)
- Perspective (For whom?)
- Intervention (What?)
- Comparison/counterfactual (Compared with what?)
- Evaluation (With what result?)

34. Additional information was extracted from each evaluation regarding level of intervention using the conceptual framework; evaluation and methodological details; education initiative description; effectiveness and impact results; other evaluation findings; and lessons learned and policy recommendations. The data extraction framework was pilot-tested and modified to ensure the major features of the evaluation dataset were captured.
35. **Data analysis and synthesis:** An iterative approach was used for data analysis and synthesis, which started by grouping evaluations by key 'levels of intervention'. In addition, evaluations were grouped by target population. Key themes and concepts emerging from the evaluations were identified and highlighted during this process, using an inductive procedure. This stage of the review was highly iterative, collaborative and time intensive, requiring the review team to constantly move between the data, emerging themes and findings (Langlois, Daniels & Akl 2018).
36. **Assessing evidence of success:** Unlike syntheses of impact evaluations (which are experimentally designed and examine discrete interventions), evaluations in this dataset were frequently of complex, multi-component interventions. In these evaluations, qualitative and process tracing methods were often used to provide rich contextual information that address questions about how complex interventions function, why, for whom and in what contexts. Throughout the analysis, when assessing and reporting evidence of success for educational outcomes, the focus was, first and foremost, on the subset of evaluations that included impact evaluation methods, such as a comparison/control groups or appropriate statistical analysis (e.g. propensity scores, difference-in-difference estimation), in order to assess causal effects and attribution of the initiatives under evaluation.
37. For evaluations that use qualitative methods, it is not possible to attribute quantified effects to an intervention. However, these evaluations highlight contextual information, encompassing factors such as relevance and sustainability of initiatives under evaluation, and barriers and facilitators to implementation of these initiatives in education systems. The review also focused on challenges faced when implementing complex, multi-component interventions. As noted later, evidence of this kind is largely missing in impact evaluations and other experimental designs. Such evidence can be used to point to promising areas of intervention that deserve further rigorous evaluation and can provide information about the organisational effectiveness of the development actors who sponsor or implement specific interventions with a gender parity, equality and inclusion focus.
38. **Peer review and oversight:** The commissioning institution, UNESCO's Internal Oversight Service (IOS) was responsible for overseeing this study, and provided peer review and guidance regarding the scope, methodology and approach for the study. A reference group led by IOS and comprised of representatives of six organisations was formed to play an advisory role in relation to the initiative. The organisations represented included: Education Cannot Wait, Norad, UNESCO, UNICEF, the World Bank and the World Food Programme. This reference group helped to finalise the study approach and the terms set for inclusion or exclusion of evaluations. They also reviewed and advised on initial findings and drafts of the report.
39. **Limitations:** The evaluations collected for this evidence synthesis are diverse: they ask different types of evaluation questions and report on projects, programmes and activities using varied methodologies and indicators. As noted above, unlike the evidence maps and systematic reviews of experimental studies promoted through the work of other organisations, evaluations in this set rarely report on impact or effect sizes using a standard impact evaluation methodology (Government of Australia, 2017; 3ie, 2019; Sniltuveit et al., 2016). Furthermore, the evaluations in this set asked different types of questions; used different concepts of inequality; and used a variety of measures to report on outcomes. This variation made the synthesis of evaluation findings on outcomes and impact complex and challenging. Further limitations were imposed by the study timeframe and resources:

- Resource and time limitations led to the decision to select independent evaluations from a limited number of UN and other organisations, and to exclude research evidence published by research and other units of each organisation (including impact research studies), as well as mid-term and self-evaluations. The 'cut off' time frame for the evidence collected was limited (four years), particularly when compared to other rigorous reviews which often synthesise findings over 10 years or more.
 - The team was not commissioned to systematically situate or compare the report findings to the wider research evidence on equality and inclusion. The study does not include research on the evaluation practices and policies of included organisations and therefore does not report in-depth on how organisations select, prioritise and use their evaluations.
 - The team was not commissioned to appraise the methodological quality of individual evaluations.³
40. Despite these limitations, the findings presented in this report are robust and map the current landscape of the evaluations in the field of education commissioned by 13 organisations during the period reviewed. This review has unique value, especially for identifying key evidence gaps, and will serve to complement other sources of synthesised evidence.

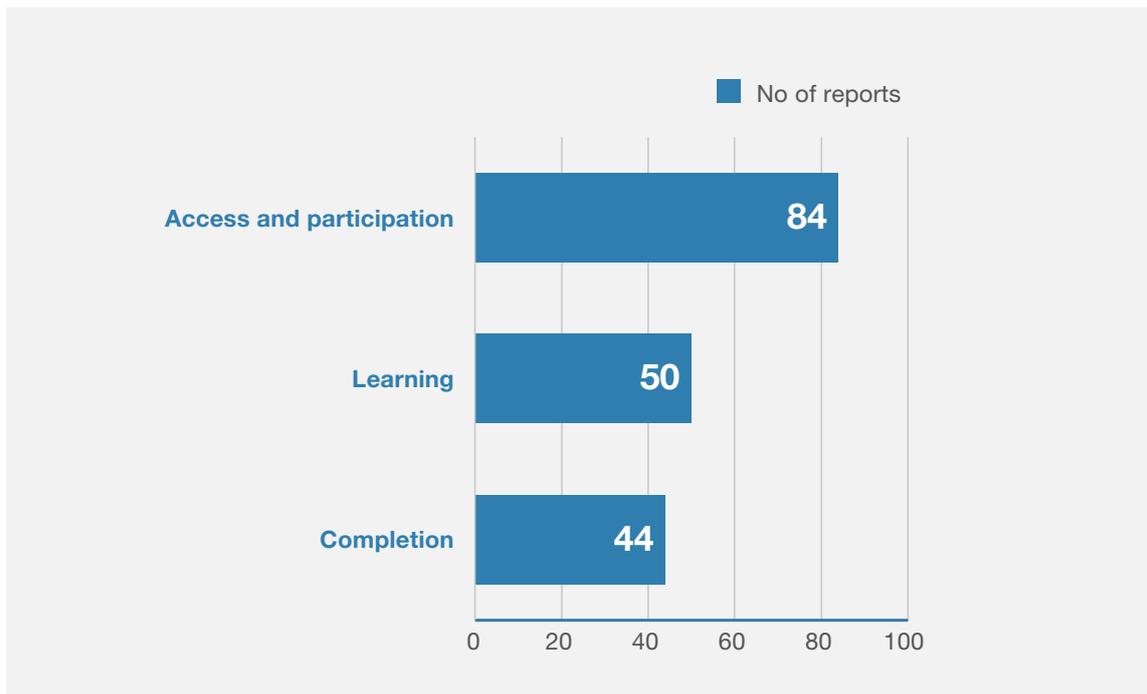
³ This type of quality review was not included in the scope of work for this report but is supported in other similar synthesis efforts – see for example USAID's recent synthesis of its education sector evaluations and its new evaluation quality assurance tool and a DFID funded synthesis of evaluations of landmine action (USAID, 2017, 2018; IOD Parc, 2012).

5. Evaluations by geography, level of analysis, questions and methodologies

Question 1. (a) To what extent is SDG 4 Target 5 covered by the publicly available evaluations undertaken worldwide by major multilateral and bilateral aid agencies since 2015? (b) Among those covering SDG 4 Target 5, what types of questions do the evaluations seek to answer and using what methodologies?

41. This section provides a descriptive mapping of the evaluations of education initiatives that include equality and inclusion-related objectives or results. The section starts by describing how SDG 4 Target 5 is covered within the evaluation dataset, and then describes the distribution of the evaluations by type of result(s) reported (i.e., access, survival or learning) by donor organisation and by level of education. The section concludes with an analysis of the questions asked and methodologies used in the evaluations.
42. **The vast majority of education initiatives evaluated objectives or outcomes related to SDG 4 Target 5.** Only 9 of the 156 identified evaluations of education initiatives in the original search were excluded because they did not report on objectives or outcomes related to gender parity, equality and inclusion. This is good news: even in evaluations of interventions that primarily focus on education quality, learning outcomes and building system capacity, attention to equality and inclusion is the norm. While few of the evaluations make an explicit reference to SDG 4 Target 5, other SDG 4 targets are mentioned (e.g., early childhood education and development, vocational education) and links to SDG 4 Target 5 can be inferred in all.
43. **Improved access to education was more frequently reported as an outcome than survival and learning.** As shown in Figure 5-1 below, evaluations reported less frequently on equality and inclusion-related results in terms of completion and learning outcomes than on access. Some evaluations cite other types of gender and equity-relevant outputs and outcomes, such as strengthening gender-responsive and inclusive teaching practices, improving employability and life skills of marginalised groups and reducing gender violence in schools. The measurement of equality and inclusion was often not the primary focus of evaluations in measuring the impact of interventions. Similarly, data gathered on the effects of education initiatives in terms of equality and inclusion tended to be unsystematic and limited to comparisons between groups.

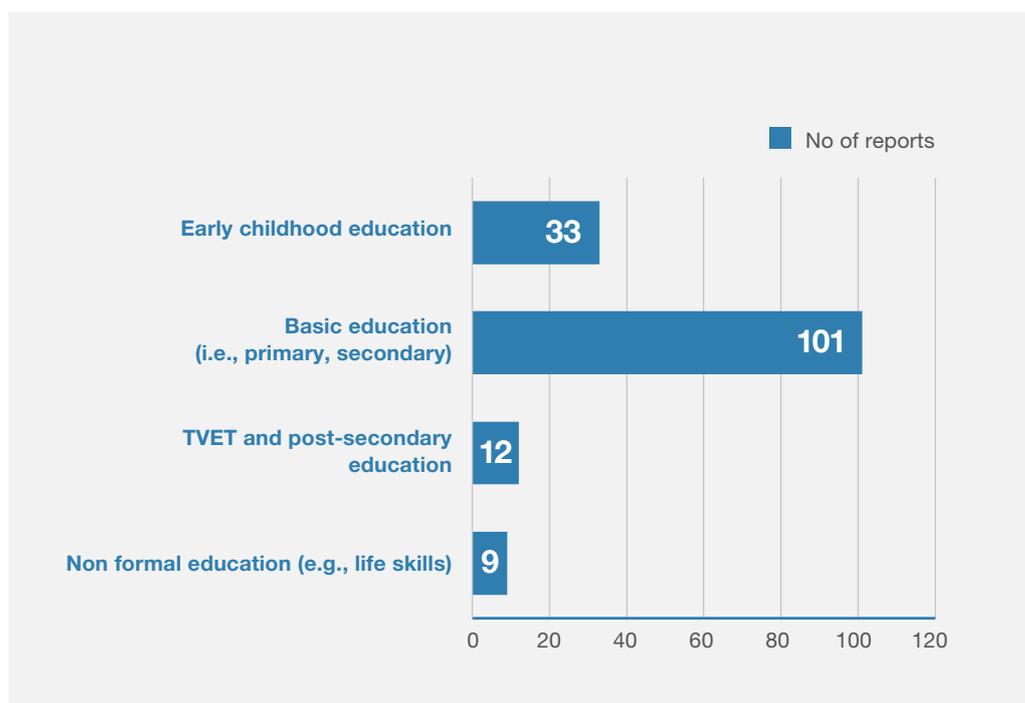
Figure 5-1: Outputs/outcomes reported in the evaluations



Distribution of evaluations by level of education

44. **A majority of evaluations are of interventions in basic education** (i.e. primary and/or secondary education), reflecting the focus of the international education community since the launch of the EFA movement in 1990. As shown in Figure 5-2 below, more than 30 evaluations include a focus on **early childhood education**, typically initiatives where pre-primary education is attached to primary schools (6 of these were stand alone, while 27 were combined with other education levels). These early years' initiatives focus on improvements in equality and inclusion for marginalised groups such as children with a disability and Roma children and those in refugee settings.
45. In contrast, evaluation evidence related to tertiary education is scarce. This gap underscores the need for more equality and inclusion-focused evaluations at the **tertiary level**, especially in light of global data showing that many developing countries are grappling to expand this level of education equitably. The lack of evaluation evidence may be tied to the absence of an explicit strategy for equity and inclusion in post-secondary education among development organisations.

Figure 5-2: Level of education covered by evaluated initiatives⁴

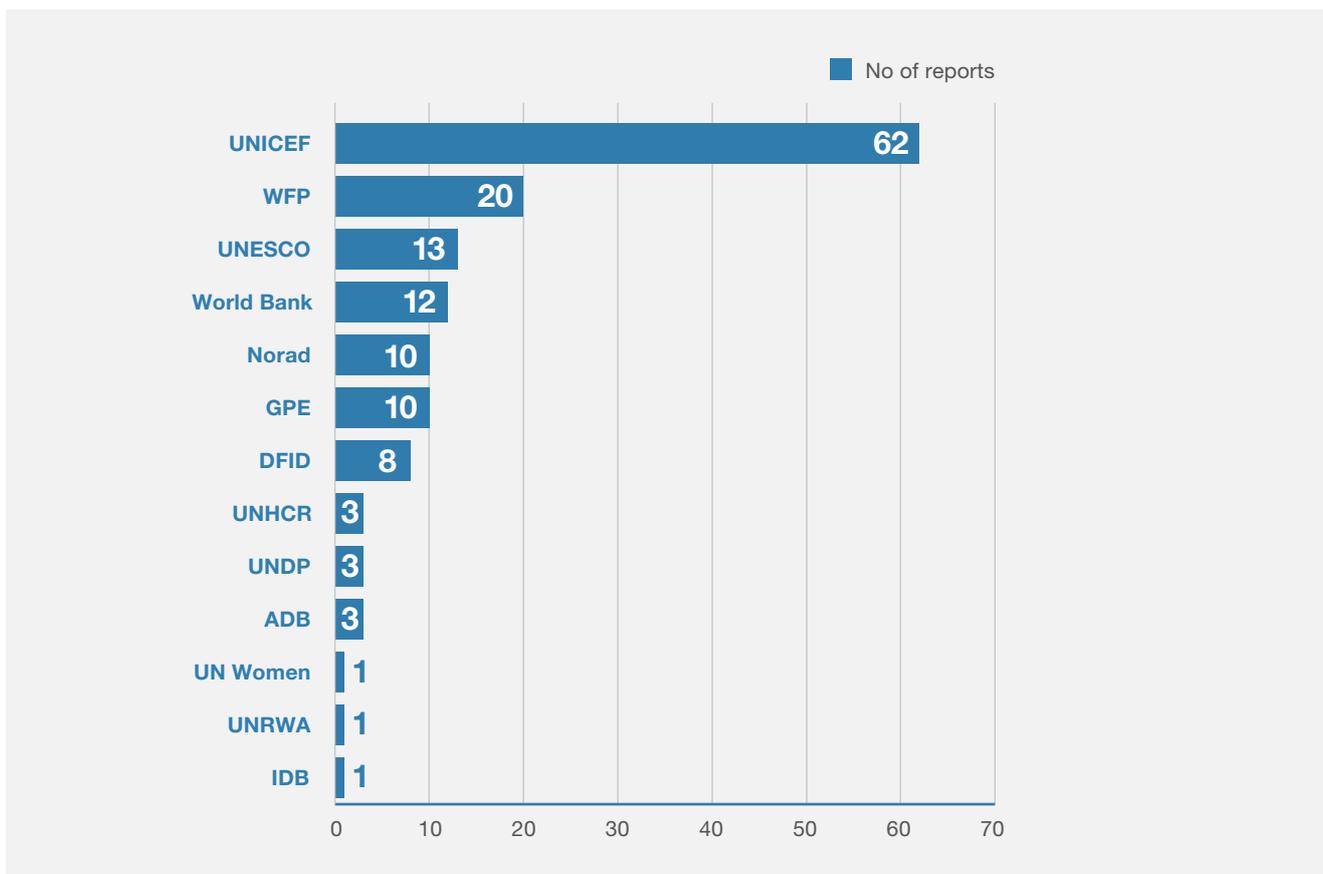


Distribution of evaluations by organisation

46. A total of between 30 and 40 evaluations were produced annually by all organisations in the review combined. There is considerable variation in the number of evaluations contributed by each organisation, as shown in Figure 5-3, as well as in terms of the intervention levels that they report on. UNICEF contributed the largest number of publicly available evaluations related to educational equality and inclusion, followed by WFP. The volume reflects the type of initiatives being sponsored by these organisations, which are often of discrete interventions. For example, the WFP regularly evaluates its school feeding programmes across different countries. UNICEF, which is highly decentralised, regularly publishes evaluations by its regional and country offices on its website, while UNESCO does not, leading to a smaller overall number contributed by UNESCO. UNICEF and UNESCO contributed heavily to thematic cross-national evaluations.
47. Norad, the Global Partnership for Education (GPE), the United Kingdom’s Department for International Development (DFID), and the World Bank were more likely to contribute evaluations with a strong system or sector reform focus. Such evaluations typically report on interventions operating at multiple levels of intervention. The smaller share of evaluations contributed by the World Bank (and other multilateral development banks) can be explained both by the type of interventions in their education portfolios and by their evaluation strategies. For example, World Bank initiatives tend to be complex, multi-component and large-scale. It also uses a system of self-evaluation by task managers, followed by review and validation by the Independent Evaluation Group (IEG), as a central pillar in its evaluation strategy. As noted earlier, these Implementation Completion Reports (ICRs) were excluded from the study because they did not fulfil the eligibility criterion of being an independent evaluation. Only fully-fledged Project Performance Assessment Reports (PPARs), which involved systematic evaluation by the IEG-WB, were included.

⁴ Many interventions use the term “basic education” to include a mix of primary and secondary level interventions. It was not possible to provide disaggregated distributions for these two levels. Numbers add to more than 147 as, occasionally, more than one level of education was included.

Figure 5-3: Evaluations in the dataset by organisation

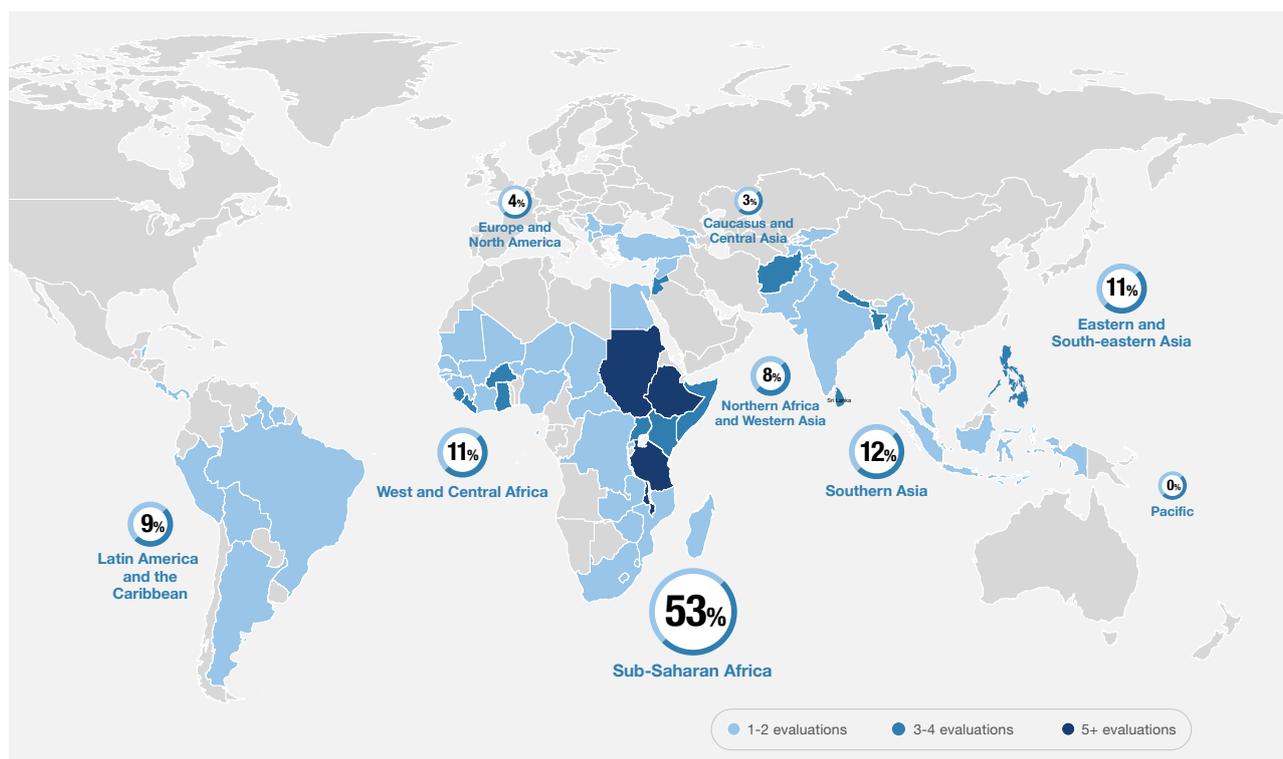


Distribution of evaluations across regions and countries

48. **The evaluations are primarily of initiatives in developing countries** (Figure 5-4). Geographically, the evaluations in the dataset were of initiatives located in 73 low- and middle-income countries.⁵ Their distribution mirrors the developing country focus of the major funding provided by most UN and multilateral agencies. The largest number of evaluations was undertaken in sub-Saharan Africa (53%). Eastern and South-Eastern Asia, South Asia, North Africa and Western Asia, and Latin America and the Caribbean each shared between 8% and 12% of evaluations. Relatively few evaluations were conducted in the Caucasus/Central Asia and Europe/North America regions. Notably, there is an absence of evaluations from the Pacific region. Ethiopia was the country with the largest number of evaluations in the dataset (10), followed by South Sudan (6), Malawi (6) and Tanzania (6). The evaluations included evidence from contexts that tend to be underreported, among them evaluations from challenging contexts and settings with Syrian refugees.

⁵ This count does not include “global/thematic” evaluations that operated across a large number of countries, unless specific country case studies are included in these evaluations.

Figure 5-4: Evaluation reports by country and region



Evaluation objectives and methodological approaches

49. **The main focus of the evaluation questions is typically the organisational effectiveness of the agency responsible for delivering the intervention**, including its ability to deliver a set of prescribed outputs. Evaluation questions are typically written with the sponsoring organisation as their main audience. However, as noted in the Introduction, the evaluations in this dataset report on a heterogeneous group of programmes and activities, situated at various levels of intervention. Not surprisingly, evaluation questions were equally varied and diverse across this set.
50. **The evaluation questions do not appear to be centrally concerned with contributing to or building shared knowledge.** It was rare to see questions aimed at assessing the contributions of the evaluation itself to the wider evidence-base for specific intervention types. Wider research evidence, including evidence from increasingly available randomised control trials of specific interventions was infrequently used to anchor evaluative analyses in these evaluations. A good practice demonstrated in some organisations was the preparation of regular evaluation syntheses or updates that aim to synthesise evidence from one organisation’s education evaluation portfolio (IOD Parc, 2012; UNESCO, 2017; USAID, 2018). Such syntheses can be even more valuable when they include research evidence and explore how corporate evaluations contribute to this larger body of knowledge.
51. Evaluations did not consistently or systematically report on all five of the OECD-DAC evaluation criteria (OECD DAC, 1991). There was considerable variation in the extent and depth of reporting on relevance, effectiveness, efficiency, impact and sustainability:
 - Relevance, effectiveness and sustainability questions were typically well-reported on, often with alignment to local context, policies and goals (although limited specification in the evaluations of relevance to whom). Sustainability was typically assessed in terms of funding continuity, not in terms of whether results are potentially transformative. Evidence in these domains was rich in contextual information, fundamental for policy planning and development.
 - Efficiency questions were weakly answered. Data on how economically inputs or costs were transformed

into results or benefits as measured, for example, by cost-benefit ratios, was rare. Evaluations in the dataset often used qualitative data to fill this gap, for instance, asking key informants whether activities were implemented in the most efficient way compared to alternatives.

- Impact was weakly evaluated. As described below, relatively few evaluations included the use of a counterfactual or a theory-based design against which the project or programme’s impact could be measured. Furthermore, a large number of evaluations noted that the monitoring data available were insufficient or not reliable.

52. **Evaluations typically adopted a pragmatic and descriptive approach, primarily utilising qualitative methods.** Most involved triangulating information from desk reviews of project reports and available monitoring and evaluation documents, stakeholder interviews, field visits and secondary analysis of national, regional or school-level education data. Many evaluations were carried out in crowded spaces where multiple development projects were taking place, making attribution difficult. These were, by and large, contribution-based evaluation methodologies (Mayne, 2008).
53. Quantitative methods that developed a counterfactual to ascertain programmatic impact and causality were used in less than 20 percent of the evaluations, or 28 overall (Table 5-1). Of these, three evaluations included a randomised control trial (RCT) and 17 evaluations adopted quasi-experimental designs, with estimation strategies such as regression discontinuity design and propensity score-based weighting to improve causal inference. The remaining eight evaluations used natural experiment designs with a comparison/control group. In addition to those studies with quantitative counterfactuals in their designs, two evaluations in the dataset referred to an impact evaluation conducted as part of the implemented project and used data from this research to answer questions about impact.

Table 5-1: Evaluations using quantitative counterfactuals

Study design	No of reports
Randomised control trial	3
Quasi-experimental design	17
Natural experiment	8

54. These evaluations focused on assessing the impacts of precisely defined interventions, primarily cash transfers and school feeding, in terms of educational outcomes related to equity, equality or inclusion (see Box 5-1).

Box 5-1: Evaluations using quantitative counterfactuals by intervention type

- 8 cash transfer programmes
- 5 school feeding programmes
- 3 multi-component programmes
- 3 child-friendly schools
- 2 life skills programmes
- 2 results-based aid
- 2 school infrastructure
- 2 education management
- 1 alternative basic education

55. An important challenge is how to adapt evaluation methodologies to evaluate the impact of multi-component system or sector-wide education initiatives. Such large-scale initiatives are becoming ever more common as a modality for development assistance. Yet, evaluations that involved complex education initiatives often did not reflect this complexity in the nature of the evaluations and the type of impact assessed. Instead, they typically focused on delivery of discrete outputs and achievement of expected outcomes and rarely sought to explore how components fit together to achieve sustainable, system-level change. This challenge has been noted in several other recent reports on evaluation in international development settings (Dahlgren, 2007; Samoff, Leer, & Reddy, M., 2016; USAID, 2018).
56. **Theory-based methods show promise when evaluating complex, system-wide initiatives.** One effective methodology used in a number of evaluations of system-level interventions involved comparing how a specific programme or intervention worked across a number of country contexts (see, for example, GPE Evaluations 10-18)⁶. A number of evaluations that looked at system-level interventions also started by creating a programmatic theory of change and used this to explore a causal chain in order to systematically understand the contribution of the intervention (GPE Evaluations 10 to 18; Norad Evaluations 35 to 39). However, evaluations commonly reported that interventions under review were not designed around explicit theories of change and either lacked causal models or theories in the project design or had unclear result chains.
57. **The power of comparison between organisational strategies was not used as a methodological lever as widely as it could have been.** Only one evaluation in the dataset, 'Evaluation of Norwegian Multilateral Support to Basic Education' (Norad Evaluation 35), explicitly set out to compare the effectiveness of the education equity strategies of UNICEF and GPE. This evaluation combined an explicit theory of change and rigorous qualitative country case studies.
58. **It is rare to see experimental studies integrated into the evaluation of system-level interventions.** Few evaluations in the dataset used methodologies that compared the effectiveness of alternative approaches to solving similar challenges. An exception to this was the 'Evaluation of the Girls, Education Challenge (GEC) Fund', which explored the effects of different interventions on girls' attendance and learning (DFID Evaluation 7). However, even in this case, results were often inconclusive due to a lack of sufficient and reliable evidence linking individual activities and outputs with specific effects on education outcomes. A secondary assessment of the GEC initiative found that there was a gap between experimentation with discrete interventions and the ability to link these interventions to sustainable system-level change (ICAI, 2018). Taken together, the results of this review suggest that more can be done to embed experimental methodologies within evaluation designs that aim to understand sustainable reform of education systems.
59. **There is room for more joint evaluation approaches.** A relatively small number of jointly conducted evaluations were found in the dataset, highlighting an opportunity for donor agencies to learn and share their evaluation processes, as well as reducing the burden on recipient countries and permitting cost-sharing and broader ownership of findings.
60. **The use of participatory approaches was limited.** Evaluations in the dataset commonly reported using participatory approaches through the involvement of stakeholders, during data collection. However, participatory approaches often appeared to be nominal and extractive, rather than transformative. Stakeholders and beneficiaries rarely appeared to have been engaged in the design of the evaluation or selection of evaluation questions. Thus, while it is clear from the evidence collected in these evaluations that participatory approaches were beneficial to improving the accuracy and relevance of reported results and explaining contextual factors, it is less clear that the methodologies chosen advance the ownership and relevance of the evidence for local stakeholders.

⁶ Evaluations are cited by organisation and are numbered. The full list by organisation and number can be found in Appendix A.

61. Data challenges were identified as a severe methodological limitation. Gaps and quality issues are reported in the available evidence base for evaluating equality and inclusion. Evaluations tend to use monitoring and evaluation data that are largely output and activity-based (e.g. number of beneficiary children enrolled in schools as opposed to improved gender equality outcomes). The quality, availability and reliability of these data were often reported as inadequate to measure and analyse outcome-level results. This created a significant challenge when seeking to determine efficiency and impact. Furthermore, in studies of complex, system or sector-wide reform programmes, a common evaluation finding was that national Education Management Information Systems (EMIS) and other statistical platforms were not available to provide strong monitoring of outcomes. This frequently led to recommendations for greater investments in national M&E capacity. It was not unusual for evaluators to note that available data were not disaggregated beyond gender, thereby not allowing for analysis of equality and inclusion impacts beyond binary gender differences (see Box 5-2).

Box 5-2: Illustrative comments about data quality

"[The organisation] was active in supporting the full school participation of marginalised children (for example out of school children and those touched by HIV/AIDS), but it articulated no expected results or outcome indicators for this work. Even outputs in this area (e.g., improvement in percentage of children acquiring relevant HIV/AIDS knowledge and skills) could not be tracked because there were no recorded baseline data." (Norad Evaluation 38, Malawi)

"What you measure is what you get: Because of the prior emphasis on meeting the Education For All and Millennium Development Goals of access and gender parity, indicators for these aspects have been developed. Data were regularly collected and managed by the EMIS. However the goals of increased quality and equity have not been well measured." (Norad Evaluation 39, Nepal)

"It was not possible to monitor equity in the detail that was originally proposed... given lack of a comparison group and lack of baseline trends, attribution of outcomes to the project is difficult." (IEG-WB Evaluation 28, Indonesia)

"...there is no pre-post or comparison to allow for causal implications of the programme especially as related to equity in learning outcomes.... The equity outcomes were not well monitored." (IEG-WB Evaluation 30, Lao PDR)

6. Types of educational initiatives and target populations in the evaluations

62. Responding to Question 2, this section looks at the distribution of the evaluations by levels of intervention: (i) child, household and community; (ii) teachers and schools (including pedagogy); (iii) education systems; and (iv) global/thematic. It also explores the type of initiatives that are included under each level of intervention and the distribution of evaluated interventions by their target populations.

Distribution of evaluations by intervention level and across intervention areas

63. The majority of evaluations in the dataset are of interventions at the level of children, households and communities (Figure 6-1) and focus on school feeding, cash transfers and non-formal education (n = 46). This reflects the fact that WFP and UNICEF provided the largest number of evaluations overall, and often conduct initiatives at these levels. There is also a substantial pool of evaluations of complex, multi-component interventions that aim to improve education systems or subsectors (n = 44). Twenty-seven evaluations are of initiatives at the global level, including evaluations of an organisations' portfolio of activities.

64. **In contrast, the smallest share of evaluations is of discrete initiatives at the schools and teachers level.** Interventions at this level include programmes of support for teacher training, school and classroom infrastructure; programmes that aim to eliminate school-related gender-based violence; as well as programmes that provide integrated packages of teacher training, materials and coaching (as for example in the case of programmes that promote 'child-friendly schools'). Interventions related to teachers and schools are also embedded in many system-level programmes evaluated, but the evaluative evidence provided on them is limited.

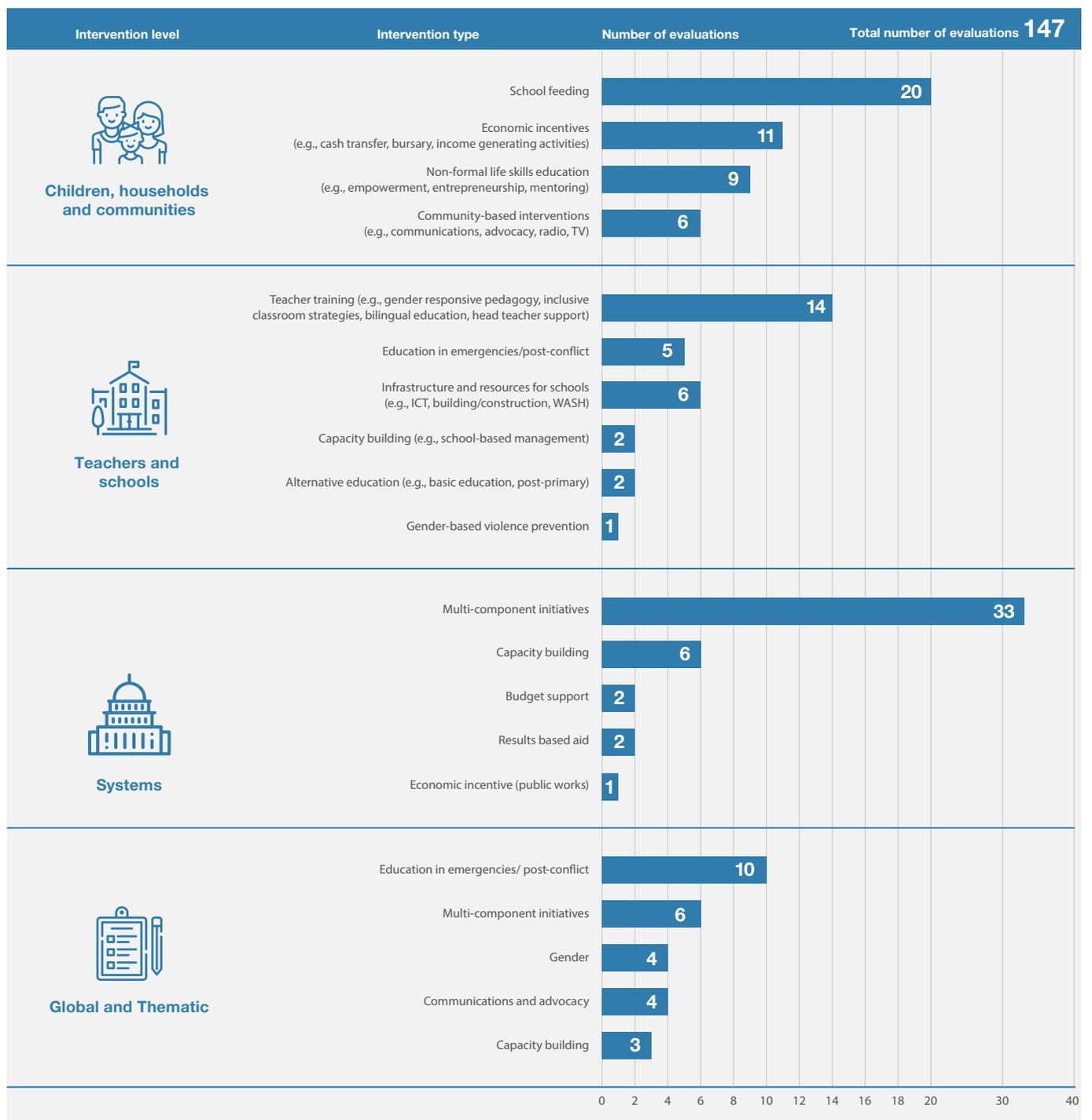
65. **Forty-six evaluations are of interventions at the level of children, households and communities.** Most of these evaluations were of school feeding, followed by economic incentives such as cash transfers and non-formal life skills initiatives focused on girls' empowerment, entrepreneurship skills, mentoring and job shadowing and adult literacy. The following is a broad description of these initiatives and how they promote gender parity, equality and inclusion for vulnerable groups:

- **School feeding initiatives** typically include, as a core objective, increased access to pre-primary and primary education. School meals constitute a conditional safety net while promoting other benefits, such as improved nutrition, attendance and attentiveness for marginalised children. Gender equality is often integrated into these initiatives, for example, in contexts that lack gender parity in school enrolment, projects might provide a monthly take-home ration to girls as an incentive to attend school. School feeding initiatives are also implemented in refugee and other emergency contexts to help support achievement of education improvement targets and improved nutrition.
- **Economic incentives** include conditional or unconditional cash transfers for households living in poverty, and in some cases girls, and typically have as an objective improved enrolment and attendance in primary or secondary school through a cash benefit, thereby reducing financial barriers to education. Other initiatives include bursaries to post-secondary education and community development/social protection funds to both create employment and build schools. Economic incentive initiatives tend to focus on access outcomes in their objectives. The main causal mechanism through which cash transfers affected educational outcomes included: decrease of the direct costs of schooling (textbooks, uniforms) and indirect costs (loss of income).
- **Non-formal life skills education** initiatives conducted in the community were diverse in their objectives, often in conflict-affected and fragile contexts, and focused on literacy training, vocational skills, conflict resolution and peace-building and mentoring for adolescent girls, among others. A unique aspect of these initiatives was their focus on beneficiaries' psycho-social well-being as a measured outcome.
- In addition, there were some examples of community-based interventions such as communications

and advocacy programmes that sought to change social norms or mobilise communities to invest in education for girls, and radio programming that sought to enhance parental engagement and promotion of early learning.

66. **Thirty evaluations were of interventions targeting teachers, teaching and schools.** Interventions evaluated included programmes of support for teacher training for Child-Friendly Schools, bilingual education, gender responsive and inclusive pedagogy and multi-grade teaching; infrastructure and resources for schools such as online learning and distance education, school and classroom construction and provision of water, sanitation and hygiene (WASH) facilities; education in emergencies interventions; and alternative, flexible education initiatives especially for those in rural and remote contexts. Further details about these programmes are provided below:
- **Teacher training initiatives** were diverse, sometimes focusing on gender-sensitive and inclusive teaching practices, in order to provide a non-discriminatory learning environment. Direct links to school enrolment and completion for girls or vulnerable groups were rarely made, but objectives and results tended to focus on changes in teacher practices and improved learning environments.
 - **Infrastructure and resources for schools** included building and rehabilitating schools and classrooms to promote school enrolment, in addition to the provision of WASH facilities, with an objective of promoting menstrual hygiene management and improving girls' enrolment and attendance in school. Some initiatives included information and communication technologies (ICT) to facilitate distance and open learning opportunities for youth in rural areas; these initiatives focused on secondary, technical and vocational education and training (TVET) and higher education. They improve access to schooling by increasing the availability of schools and reducing distance and travel time. New or refurbished schools attract children. By improving the learning environment, students' learning experiences may improve. Provision of latrines may be important for ensuring the retention of girls in schools.
 - **Education in emergencies/post-conflict** initiatives addressed immediate education needs and enabled children to enrol in school, through activities including building temporary learning shelters, schools and teacher accommodation, the provision of textbooks and learning materials, recruitment of teachers, teacher training and government coordination.
67. **Forty-four evaluations were of system- or sector-level interventions** which aimed to improve education systems or sub-sectors by investing at multiple levels, including in schools, teachers, learning materials and curriculum, as well as national policies and institutional capacities. This category included sector-wide education programmes that provided results-based aid, budget support and capacity building support. Many of these evaluations included components similar to those described in the Snilstveit et al. study (2015, 2016) as system-level intervention – for example, school-based management or decentralisation reforms, teacher management/allocation, economic incentives and public-private partnerships, as well as improvements in education management information systems, research, policies and legislative frameworks.
68. **Twenty-seven global and thematic evaluations were included in the sample.** These evaluations were of diverse interventions often looking at a cross-national portfolio of activities organised around a specific topic or theme, for example, support to higher education, gender equality in education capacity building or support to children affected by conflict and crisis. Three evaluations at the global level focus on advocacy and communication: the evaluation of the Global Education Monitoring (GEM) Report, UNICEF's Communications for Development programme and GPE's Civil Society Education Fund. These evaluations typically ought to explore the efficacy of organisational priorities and strategies and, in several cases, include country case studies to explore how a variety of contexts shape implementation and outcomes.

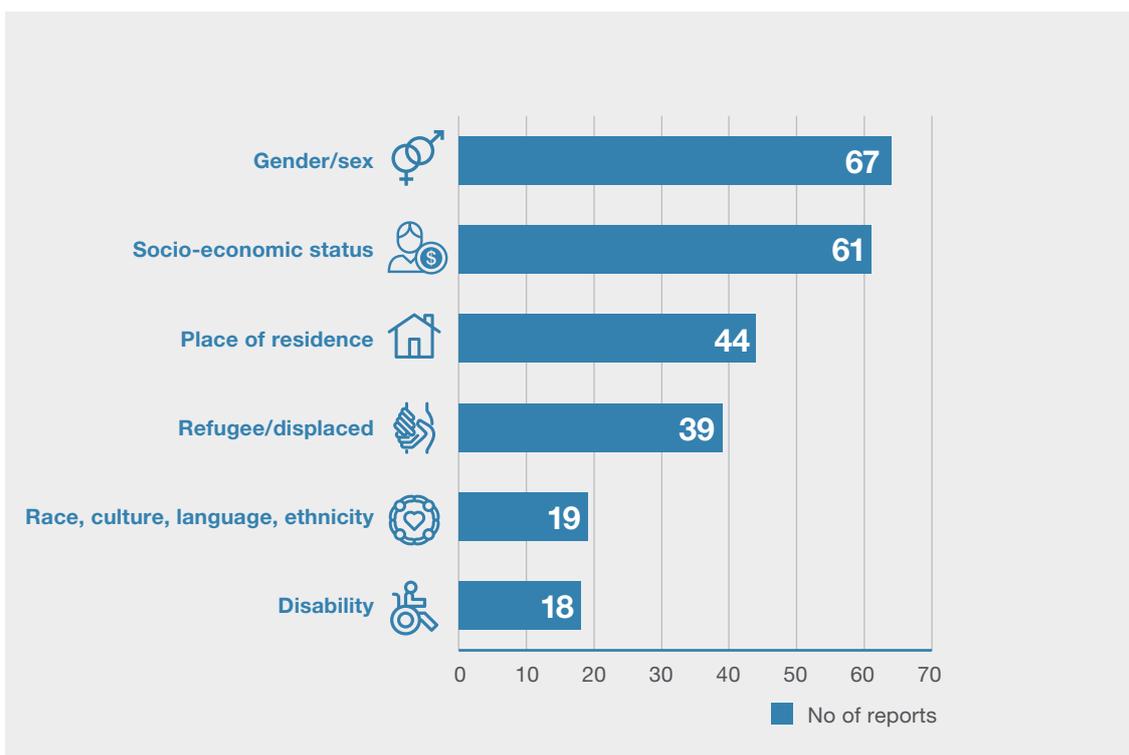
Figure 6-1: Distribution of evaluations by intervention level and across intervention areas



Target populations addressed in the initiatives under evaluation

69. **When looking at the populations targeted in the interventions evaluated, gender/sex emerges as most commonly addressed in the evaluated initiatives (Figure 6-2).** Initiatives often targeted girls and were sometimes exclusively designed to support girls' education. A limited number of initiatives were specifically targeted towards boys (e.g., UNICEF Evaluation 111), even though several evaluations noted that in some contexts more support is needed for boys, especially to reduce drop-out in secondary school. Only one evaluation involved an initiative that focused on sexual orientation and gender identity including target populations such as lesbian, gay, bisexual or transgender. The predominant focus on gender/girls in these initiatives is perhaps not surprising given that these evaluations report on programmes carried out during the era of the MDGs, which had specific targets for girls' education.
70. **Socio-economic status and place of residence (urban or rural, sub-national region) are common areas for targeting** across the evaluated interventions (n = 61 and 44, respectively). Location is a dimension of inequality that appears to be relatively easy to measure and document. Several initiatives also targeted households living in poverty. An overall concern across these evaluations, however, is the availability of good quality disaggregated data to allow for evaluation of the impact of interventions on populations facing intersecting inequalities.
71. **A unique aspect of this dataset is the large number of evaluations of interventions that target children in fragile, conflicted-affected, displaced and refugee settings.** As noted earlier, a large number of the interventions evaluated have been implemented in conflict-affected settings and specifically target the needs of displaced children. A unique feature of this evaluation dataset is that it provides insights into equity in such settings that are often unavailable.
72. **There is a serious lack of evaluations of interventions that target the needs of children with disability and other dimensions of equity,** such as marginalisation by race, culture, language and ethnicity. For these populations, evaluation evidence is very limited even though these dimensions are widely recognised as factors that are likely to create educational disadvantage.

Figure 6-2: Disadvantaged populations targeted by educational interventions



7. Lessons from the evaluations about what works to support SDG 4 TARGET 5

Question 3

What evidence of success is found by the evaluations regarding initiatives that tackle gender disparity and exclusion? Where do the evaluations point to initiatives that have the potential for replication and what policy actions do the evaluators recommend?

Question 4

What do these evaluations tell us about key contextual factors and how they interact with these initiatives? What are the barriers and facilitators to effective implementation of these initiatives?

73. This section looks at findings and lessons learned from the evaluation dataset in order to answer Questions 3 and 4. As described in the methodology section, a narrative approach was used to synthesise evaluation findings, recognising that the absence of standardised measures and counterfactual evidence does not allow for robust attribution of outcomes or effect sizes.

Lessons from evaluations of interventions: Children, households and community level

74. **Interventions at the children, households and community levels included the largest number of evaluations with a counterfactual**, allowing for a robust assessment of promising areas for addressing equality and inclusion. Such interventions primarily addressed barriers and constraints faced by households. Evaluation findings can be synthesised by programme type.
75. **Economic incentives/cash transfer programmes.** The review included 11 evaluations of cash transfer programmes and an evaluation of a bursary initiative not including the many sub-evaluations reported in DFID's Girls, Education Challenge Initiative (DFID Evaluations 2-4), with most evaluations including a comparison or control group. The initiatives took place in 10 countries, mostly in Eastern and Southern Africa and Latin America, and targeted families with children in pre-primary, primary and/or secondary school.

Box 7-1: Cash transfers have a strong impact, but some limitations

In this dataset, cash transfers are shown to have strong impacts on school attendance (11 studies). Cash transfers are among the most widely and most studied of any international development intervention – representing the largest number (50) of impact evaluations in Sniltsevit et al.'s (2015 & 2016) systematic review.

76. **Evaluations based on experimental designs demonstrate that cash transfer programmes have positive impacts on school participation among children from poor and vulnerable households.** As is the case in the wider evidence base captured in systematic reviews (Bauchet et al., ; 2018; 3iE, 2010 J-Pal, 2017; Sniltsevit et al., 2015 & 2016), conditional cash transfers improved school attendance among children in poor households in Tanzania (UNICEF Evaluation 104), Malawi (UNICEF Evaluation 65), Liberia (UNICEF Evaluation 110) and Ethiopia (UNICEF Evaluation 69), as did unconditional transfers in Nigeria (UNICEF Evaluation 73). They can serve to boost enrolment for girls in schools where gender parity has not been achieved (UNICEF Evaluation 73) and can reduce children's work and household chores (UNICEF Evaluation 80). Bursaries can lead to improved access to post-secondary education for vulnerable youth and refugees as in Jordan (UNESCO Evaluation 49).

77. **Some qualified constraints on the effectiveness of cash transfers are detailed in this evaluation set and deserve further exploration.** The effects of cash transfers on school completion and dropout rates are inconclusive, with at least one evaluation reporting that drop-out rates in the transition to secondary school continue in spite of cash transfers (UNICEF Evaluation 98). Cash transfers will not work where there is limited supply of schooling (as in a Syrian refugee programme, UNICEF Evaluation 80). One evaluation indicated a positive impact on learning outcomes, likely due to increased time for study (Tanzania – UNICEF Evaluation 104). On the other hand, qualitative data suggest that cash transfers can bring bullying and jealousy towards beneficiaries. They may also negatively impact the quality of schooling through sudden increases in enrolment (UNICEF Evaluation 73). Questions about their long-term sustainability, once donor funding ends, were raised in several evaluations.
78. **School feeding programmes have strong positive effects on primary school enrolments,** particularly in areas of high food insecurity including drought-affected and internally displaced/refugee contexts, based on the 20 evaluations of interventions in our dataset, including five with a comparison control group. School feeding programmes can be particularly impactful for girls and internally displaced populations, including in Ethiopia (WFP Evaluation 146), Kenya (WFP Evaluation 133) and South Sudan (WFP Evaluation 140). As with cash transfers, school feeding programmes appear to have less conclusive effects where children are already in school (Cambodia, WFP Evaluation 134; Lao PDR, WFP Evaluation 136). The evaluations in the dataset provide some positive evidence to support the effects of school feeding on improvements in school completion and drop-out rates, in keeping with an emerging evidence base that points to strong causal links between school feeding and retention (see for example, Aurino, Tranchant, Sekou Diallo, & Gelli, 2018; Tranchant, Gelli, Bliznashkab, Sekou Diallo, Sacko, Assima, Masset, 2018; Gelli, Masset, Folson, Kusi, Arhinful, Asante, Drake, 2016). However, several evaluations note that high primary school drop-out rates remain of concern and often relate to employment for boys and early marriage for girls.
79. **This dataset is inconclusive about the effects of school feeding effects on learning outcomes,** in contrast to more hopeful outcomes found in some other impact evaluations and systematic reviews (Aurino et al., 2018, Bashir, Lockheed, Ninan and Tan, 2018, p. 180; Bundy, de Silva, Patton and Horton, 2018; Snilstveit et al., 2015 ; Gelli et al., 2016). There is a need for more regular collection of learning outcome data in this area.
80. **Sustainability and unintended outcomes are noted challenges in evaluations of school feeding programmes.** Unintended results identified include overcrowded classrooms, sometimes aggravated by the conversion of classrooms into food storage and eating places and additional demands on teaching staff and school management, which may detract from their core educational roles. Sustainability of externally sourced foods is also a challenge to which organisations have responded through the piloting of new approaches, including local sourcing of food and use of cash transfers (i.e., Kenya, WFP Evaluation 133). The WFP also reports that it is exploring the cost-effectiveness of school feeding programmes.
81. **Non-formal life skills interventions.** The review included nine evaluations of non-formal education life skills initiatives, several of which were directed at adolescent youth and young adults in conflict-affected contexts. In addition, several system interventions included investments in alternative education, among them one of an adult education initiative. These included novel initiatives aimed at changing youth behaviours, attitudes or engagement in Somalia (UNICEF Evaluation 87) and South Sudan (UNICEF Evaluation 113). There is a real need for further experimental evidence on them in order to test their efficacy for further scaling. Non-formal life skills interventions are reported as having the potential to improve transitions from school-to-work, entrepreneurship and employability skills and psycho-social outcomes including confidence and empowerment.

Box 7-2: Better evidence on non-formal education is needed

Alternative education, life skills education and programmes for accelerated learning are frequently evaluated by international organisations, typically using poor quality data on outcomes and impacts. There is a clear need for better evidence on what works and why for this type of intervention.

82. **Community-based information and advocacy interventions are evaluated in 6** discrete initiatives specifically focused on behaviour and social change among the wider community, while four strategies (i.e., UNICEF Communications for Development - C4D) focused on provision of education-related information to community members including parents and teachers. The interventions primarily focused on cross-cutting issues such as participation and gender equality. Several global/thematic level interventions also included communications and advocacy components – such as a multi-country programme by UNICEF to support communication for development, which found inconclusive effects from its educational components (UNICEF Evaluations 92, 117-119). Qualitative evidence provided of community-based activities, such as radio programming, community workshops and mothers' groups, indicated they hold promise for improving attitudes towards girls' education (e.g., DFID Evaluation 4); however, as indicated in other systematic reviews, there is an overall lack of consistent evidence of impact in this area (Snilstveit et al., 2016).

Lessons from evaluations of interventions: Teachers and schools level

83. **Evaluations of interventions that specifically targeted teachers and schools comprised a small share** of evaluations in the overall dataset, and only 7 evaluations of interventions at this level include a comparison/counterfactual. In general, they tend to provide data regarding the timeliness and quality of inputs into education processes (e.g., number of schools constructed, types of training delivered), rather than impacts on learning and other child-level outcomes. Few evaluations in this dataset look at interventions that use incentives or rewards to shape the behaviours of teachers, an area that has inconsistent outcomes according to the wider impact evaluation literature (Evans & Popova, 2016; Snilstveit, 2016, p. 34).
84. **Most evaluations do not demonstrate strong causal linkages between teacher development or other pedagogical inputs and more equitable distribution of learning outcomes.** The evaluation dataset included 14 evaluations of teacher training initiatives, as well as numerous evaluations of system-level interventions that included teacher training and other pedagogical components. However, the evaluations in this dataset are not able to tell us with any rigour whether the evaluated approaches improve equality in student enrolment or learning, as this type of data is rarely collected. There is some evidence to suggest that improved teacher training increases school completion rates, especially for girls, in Burkina Faso (UNICEF Evaluation 75). Most evaluations focus on intermediary factors, such as observations of how often girls are asked questions by teachers (compared to boys) and teachers' identification and inclusion of children with additional learning needs in their question-asking practices (UNICEF Evaluation 91).
85. **Many of the evaluations were of interventions focused on training teachers to use inclusive or child-centered pedagogies, which evaluators often reported as promising** (e.g. UNICEF Evaluations 75, 91). This is in keeping with a growing literature that highlights the impact of structured pedagogical programmes on learning outcomes, emphasising the value of coaching teachers to tailor their instruction to the learning levels and needs of students (Evans & Popova, 2016 ; Popova et al., 2018; Snilstveit et al., 2016).
86. **There is an urgent need for future evaluations that compare the equity and equality impacts and cost effectiveness of different approaches to pre-service and in-service teacher training and other packages of pedagogical inputs,** in part because both equity-focused disaggregation and cost data is missing both from this dataset and in many impact evaluations of teaching and learning interventions. Evaluations in this area rarely include measures of impact on learning outcomes.
87. **Most evaluations of infrastructure interventions do not provide good quality disaggregated data about their impact on equity.** In system-level interventions, evaluative evidence typically focuses on showing whether the delivery organisation is effective at ensuring that infrastructure is built or delivered (or rehabilitated). The review included six evaluations of the provision of infrastructure in the form of ICT, construction or rehabilitation of school buildings/classrooms and WASH facilities. Most system-level interventions also included intervention components in these areas. The evaluation evidence in this dataset suggests that school construction has positive outcomes in areas where there is a lack of available

school buildings, for example, in providing access to education in conflict-affected areas in East Jerusalem and Gaza (UNDP Evaluation 58; UNICEF Evaluation 96). Programmes for building new schools reduced school drop-out and improved perceptions of the quality of school in Malawi (UNICEF Evaluation 112), while the provision of water, sanitation and hygiene (WASH) facilities increased school enrolment in arid, pastoralist areas in Kenya (UNICEF Evaluation 84).

88. **Results pertaining to information and communication technology programmes (ICT) and specifically online learning, report mixed results**, consistent with findings from other systematic reviews (Sniltsevit et al., 2016). In Argentina, virtual classrooms facilitated access to secondary education for adolescents and young indigenous people in isolated and difficult-to-access areas (UNICEF Evaluation 95). In Sri Lanka, an evaluation found that online post-secondary education did not improve enrolment among rural students, especially for female students (IED-ADB Evaluation 19).
89. **Interventions supporting the education of children during emergencies show considerable evidence of rapid increases in enrolments** in refugee camps/host communities following relief responses. However, challenges are identified in a number of evaluations, including transition to formal education and lack of support for adolescents. There is a widely recognised need for longer-term strategies for collaboration with host governments and systems and for deeper reflection about the sustainability of emergency education programmes (Burde, Guven, Kelcey, Lahmann & Al-Abbadi, 2015). This reflects the fact that displaced populations face longer term and sometimes permanent relocation (Dryden-Peterson, 2018).
90. **School-based management, school improvement plans, school grants and other decentralisation reforms can raise challenges for equality and inclusion.** Many of the system-level programmes evaluated included components related to school-based management, school improvement plans and school grants. While this type of intervention was rarely rigorously evaluated, core lessons about challenges encountered in delivering them emerged from the dataset. For example, an evaluation of a programme by the IEG-WB in Philippines noted that there was insufficient attention to different forms of inequality in the design of a school grants programme (IEG-WB Evaluation 33). In Nepal, an evaluation noted that the school-based management rested on untested evidence and did not appropriately assess the capacity needed for large scale decentralisation reform and had a theory of change that was “not convincing on equity links to school-based management” (IEG-WB Evaluation 27). A GPE evaluation in Sierra Leone suggested limited impact of performance-based school grants (GPE Evaluation 15), while in Zimbabwe, a school grants programme was described as “overly ambitious” resulting in schools being overwhelmed and losing a key focus on equity (UNICEF Evaluation 109). Other evaluations are more hopeful about school improvement planning especially linked to school grants (IEG-WB Evaluation 26; Norad Evaluation 35).
91. **The need for more evidence on decentralisation reforms** comes out clearly from these evaluations, mirroring the inconsistent findings about their impact on equality in terms of learning outcomes as discussed in other systematic reviews (Carr-Hill et al., 2014; Eddy-Spicer et al., 2016; Samoff et al., 2016, p. 27).

Lessons from the systems level and thematic evaluations

92. **Evaluations of system or thematic level interventions rarely included evidence that has a strong counterfactual or comparison group and they inconsistently used theory-based approaches that are commonly recommended for studying complex interventions** (Weiss, 1997; White, 2009). While many of these evaluations include components at all levels (child, community, teacher, school) evaluation evidence in the majority of system-level evaluations does not provide detailed rigorous evidence about what works. Thus, this study highlights cross-cutting lessons that emerge across two or more system and thematic level evaluations. These findings are not specific to sub-component interventions at the children, households and communities and teachers and schools levels.

93. **New types of financing are reported as having mixed success:** Several evaluations in the dataset look at the effectiveness of different forms of donor financing, including 10 evaluations that report on results-based financing (DFID Evaluations 1 and 85; GPE Evaluations 10-17). Among these, the two DFID evaluations, based on quasi-experimental designs, suggested limited impact of results-based financing on school completion, learning or gender equality and inclusion. In neither instance do they create strong incentives for governments to change their programmes or allocations, despite targeted funding for improvements in equity. These findings align with recent literature on this form of aid, which has concluded that there is no evidence that payment by results leads to fundamentally more innovation or autonomy, with the overall range of success and failure broadly similar to other aid projects (Clist, 2018). However, evaluations from the GPE, based on qualitative evidence, suggest that such interventions may improve and focus policy dialogue during the planning phase, including on matters related to equality, including equity of learning outcomes.
94. **Pooled sector financing and budget support modalities receive mixed evaluations.** These aid modalities are generally understood to promote country ownership and aid effectiveness (GPE Evaluations 10-18; Norad Evaluation 39). In Sierra Leone, findings suggested that while budget support improved overall education spending and access, it did not improve policy dialogue in areas beyond access (DFID Evaluation 5). In GPE evaluations, a common finding was that, despite a commitment to pooled approaches, in only three countries were such modalities actually used by donors.
95. **Sector plans and sector-wide programmes may be too complex.** A common finding across system-level evaluations is that the programmes are too complex and ambitious, often underestimating the time, resources and capacity needed for their implementation. This is particularly well illustrated in the GPE evaluation synthesis and Norad’s Basic Education Evaluation synthesis and case studies, both of which highlight the fact that the sector plans themselves, while improving, typically lack achievability and prioritisation and are so complex that they often lead to low rates of implementation. Other evaluations also emphasise complexity in country programme designs (Zimbabwe, UNICEF Evaluation 109; Sri Lanka, IED-ADB Evaluation 19; Nepal, IEG-WB Evaluation 27).
96. **Equality often stops at universal access and gender parity, while disability, language and ethnicity are often neglected.** One of the most intriguing findings across evaluations pertains to the definitions of, and objectives for, equity and equality within system-level interventions. While these problems may be related to the targets and objectives set in the MDG/EFA era, they highlight the opportunity and need for much more precise and differentiated goals and measures of equality in evaluations. For example, several evaluations note that definitions and measures of equity tend to equate universal access with equity, based on the assumption that improving access through mass expansion will solve key equity and equality challenges. This leads to missed opportunities to target interventions effectively to those most in need, or to measure outcomes for different disadvantaged groups (Brazil, Bulgaria, Philippines IEG-WB Evaluations 24, 31, 33; Norad Evaluations 35-39; GPE Evaluations 10, 18). See Box 7-3.

Box 7-3: Illustrative comments on equity and access objectives in system-level interventions

“Equity objectives require clear definition and measurement.... In this operation equity was stated as an objective, but the equity concern was not clearly articulated or measured... Equity was defined in terms of school-level performance; the selection/targeting process was weak and didn’t meet those most in need.” (IEG-WB Evaluation 24, Brazil)

“...the clarity of the objective to improve equity in learning outcomes had important shortcomings. Both projects implicitly mix a more traditional definition of equity - relative to improvements among certain vulnerable or disadvantaged populations - with an interpretation that focuses on access for all as an approach that will, by definition, capture the less advantaged.” (IEG-WB Evaluation 33, Philippines)

97. **Evaluations frequently point to a failure to move beyond gender parity to parity among other population groups, as a measure of equity**

(GPE Evaluations 10-18; Norad Evaluations 35-39), while others point to a failure to consider “learning outcomes” as an equity concern (Norad Evaluations 35, 38). Many evaluations at both system and thematic levels report on a lack of focus on disability and inclusion in the programme designs and in available data on outcomes. As summarised in a Norad evaluation, “while most countries have made progress in gender equality at the early school levels, this is not the case as regards the equity of linguistic minorities, disabled children and poor children” (p. 3). Good recent guidelines and studies point to the potential for common approaches to measurement and evaluation in this area (Education Equity Research Initiative, 2017a and 2017b; Rose & Alcott, 2015; UIS, 2018b and 2018c).

Box 7-4: Disability is neglected

“...the fervor put into parity of girls primary school enrollments has not yet been applied to [other] types of equity.” (Norad Evaluations 35-39)

“...limited progress was observed on learners with special needs” (GPE Evaluations 10-18).

98. **Capacity development is a common challenge.** Many system-level and thematic-level interventions include components related to building national capacity for delivering strong and equitable quality outcomes in education. Evaluations consistently point to weaknesses in capacity building components of these larger programs, while at the same time highlighting areas of good practice. For example, GPE and Norad evaluations find that sector-wide programmes do not adequately assess capacity gaps within the Ministry (GPE Evaluations #10-18; Norad Evaluations #35-39). Several other evaluations point to inadequate attention to subnational capacity needs for programmes that involve large-scale decentralisation reforms and school-based management (e.g., Ghana, UNICEF Evaluation 66; Nepal, IEG-WB Evaluation 27). Both UNICEF and DFID have created targeted national capacity development programme, and technical assistance funding is also included in many sector programme; however, this is often not designed systematically and with an eye to sustainability. These capacity development programmes are often supply-driven and do not have adequate ‘use cases’ or systematic needs assessments. They are better at building the capacity of individuals, than institutions.

99. **Evaluations repeatedly call for greater investment in equity-responsive monitoring and data at the national level.**

In particular, there is a widespread call for improving the ability of EMIS to monitor outcomes for linguistic minorities, children with disabilities and children from poor households. The importance of UNICEF-led work on Out-of-School Children, which supported countries like Nepal and Ghana in improving their monitoring of disadvantaged children is highlighted in several of the country evaluations.

Box 7-5: Out-of-School Children Initiative (OOSCI)

This initiative is noted in several evaluations as playing a cutting edge role in filling capacity gaps in the monitoring of equity, but more needs to be done to ensure country capacity to monitor equity outcomes (UNICEF Evaluation 120).

100. Room for improvement on gender equality and education: More than any other area in our dataset, gender equality emerged as a consistent focus, including in a larger number of evaluations at the systems and thematic level (UNESCO Evaluation 52; IED-ADB Evaluation 20; UNICEF Evaluation 78; OVE-IADB Evaluation 34). Despite widely acknowledged successes in closing the gender gap, these corporate portfolio evaluations highlight several common challenges:

- Organisations are sometimes opportunistic rather than systematic in their gender equality work, in part driven by funding availability. They also face challenges in balancing a focus on initiatives for girls with strategic impacts on policies and cross-portfolio gender mainstreaming. These evaluations recognise the need for more disaggregated evaluative evidence across studies at all levels and intervention types and indicate an evidence gap in evaluating the relative merits of systemic/universal interventions on girls (as versus targeted interventions). Both issues are the subject of calls for more evidence in recent evidence syntheses on gender and education (Evans & Yuan, 2019; Rose & Yorke, 2019).

- A common issue highlighted in thematic gender evaluations is the need to move forward from a focus on access, to gender equality in completion, equality at higher educational levels and equality in learning outcomes. In addition, as illustrated by the set of evaluations of DFID's Girls' Education Challenge, there are untapped opportunities for greater use of comparative and impact methodologies to better allow policy-makers to select the right mix of interventions to advance gender equality (DFID Evaluations 2,3).

101. **Global-level strategies can be better linked to country needs.** Several evaluations in the dataset report on a portfolio of interventions or a thematic strategy. Findings are quite varied from these evaluations and difficult to synthesise. What comes out clearly is the importance of ensuring that there is a strong communication with country partners about how to interpret and implement objectives set at the global level. Thematic evaluations contributed by UNICEF, UNHCR, UNESCO and NORAD on thematic work in areas like gender equality, education in emergencies, out-of-school children and others, emphasise the need for greater synthesis and sharing of lessons learned from across organisational portfolios.

Box 7-6: Greater coordination and peer learning exchange

Thematic and system-level evaluations demonstrate a clear need for greater coordination and learning across organisations – and for better coordination at the country level to ensure that evaluations support the sharing and use of evidence about SDG 4 Target 5.

102. **Read together, evaluations at system and thematic levels suggest the need for greater coordination and learning across organisations,** a finding that was echoed in the country validation workshops described in the next section. Organisations are generating important data on how to improve their own activities, which are often of relevance to the work of other development partners. Some countries, for example, Nepal, have had multiple organisations evaluating achievements in the education sector, with limited signs of coordination. Yet evaluations are not systematically planned or shared to maximise these synergies and opportunities for cumulative learning. Some organisations do provide good quality evidence synthesis from their evaluations, but these are typically used for internal learning.⁷ A rapid review of major systematic reviews and evidence synthesis suggested that evaluations in this synthesis dataset are rarely cited in the other systematic and rigorous reviews. As suggested in previous reviews of education sector evaluations conducted by international development actors (Samoff et al., 2016), and in other recent overviews of education evidence (Magrath, Aslam, Johnson, 2019), there is an enormous opportunity to coordinate better across organisations to ensure both cross-agency and country-level learning and exchange.

Cross-cutting lessons for populations affected by inequality and exclusion

103. **Gender/girls.** The number of evaluations focusing on gender equality illustrated that girls' education constitutes a significant investment in evaluation work, both symbolically and financially. The evaluations provided evidence that initiatives that reduced financial barriers or provided an incentive (e.g., school feeding, upgrading facilities) can be effective ways of encouraging school participation and attendance among girls (e.g., DFID Evaluation 4; UNICEF Evaluation 73; WFP Evaluation 132). Success in some settings required addressing the needs of marginalised girls, while at the same time responding to cultural backlash against girls for receiving priority treatment (UNICEF Evaluation 73). Interventions reviewed, though effective in promoting girls' education in primary school, were often not enough to sustain results through to secondary or higher levels of education, nor to ensure equality in learning outcomes.
104. Evaluations revealed the complexity of working on girls' education issues. Multi-dimensional barriers to girls' education include poverty, insecurity, drought and other climate-related emergencies in contexts where marginalised girls are located. Evaluations noted that these types of circumstances increase pressure to marry early or support their families with income generation. Violence and safety issues also affected girls' abilities to attend school, especially in the later grades (UNICEF Evaluations 80, 108 and 123).

⁷ For some good examples of this type of synthesis see USAID's recent synthesis of its education evaluations (USAID, 2018); the synthesis commissioned by DFID on landmine interventions (IOD Parc 2012); and Samoff et al.'s (2016) synthesis and review of evaluations commissioned by Sweden.

105. Eliminating and addressing these types of multiple and longstanding causes of inequality within short project cycles (i.e., less than 5 years) was reported as an implausible goal (DFID Evaluation 4). Another finding was that interventions lacked a cohesive and consistent approach to identifying and addressing the needs of the most marginalised girls (UNICEF Evaluation 82; DFID Evaluation 2). The complexity and context-specific nature of the challenges in girls' education make it extremely difficult to design a single solution of what works. Evaluations report that more focus is needed on promoting gender equality in enrolment and learning outcomes in the higher grades of basic education, as well as more system-wide work to promote the interests and needs of marginalised girls in national education systems and policy advocacy work, for example, addressing gender equality in educational leadership.
106. **Indigenous and other ethnic, linguistic and religious minority groups.** These initiatives can be broadly grouped into two types: those that bring educational access to minority, out-of-school children in remote areas through the provision of flexible, alternative education and virtual education and those directed at improving quality learning environments through multi-lingual or 'first-language-first' education. Effective initiatives tend to occur at the teachers-schools level, including adapting teacher training, the curriculum and how it is delivered and introducing new technology to meet the local needs of a community (UNICEF Evaluations 74 and 95). Mother-tongue bilingual education has shown evidence of effectiveness in improving learning outcomes for indigenous and ethnic minority children and youth, although causal inferences remained unqualified (UNICEF Evaluations 115, 121).
107. A common challenge identified in the evaluations was transition into formal education systems (UNICEF Evaluation 74). Furthermore, there is limited evaluation evidence of initiatives of system-level policies, programmes and support to improving indigenous and minority educational outcomes, for example, more equitable funding structures and improvements to values, norms, actions and relationships. Evaluations noted insufficient priority given at regional and national levels constrained the sustainability and full quality delivery of these interventions as mainstream education modalities for these marginalised communities. Negative perceptions of multi-lingual education among policy elites and "reactive political pressures" to maintain the dominant language as the exclusive language of learning (UNICEF Evaluation 115) point to the complexity of politics and power issues and their impact on equality and inclusion of ethnic minority students. It also reinforces the need for greater capacity-building at the government/school leadership levels.
108. **Disabled persons.** Despite the strong presence of overall equality and inclusion-related objectives in the initiatives across evaluations, they have little to say on promoting equity, equality and inclusion for disabled students. Even in those interventions that included inclusive education or disability as a target or objective, evaluators frequently noted that no evidence was provided or available to allow them to evaluate relevant outcomes. Perhaps more hopefully, evaluations argued the need for better data on children with disabilities and encouraged donor organisations to ensure that these are supported within national systems and through their own monitoring and evaluation frameworks (e.g., Norad Evaluation 35).
109. **Refugees, displaced persons and those affected by conflict.** Interventions here focused on providing economic or other inputs to remove barriers or provide incentives for children to attend school. One of the main findings from these evaluations is that efforts to remove barriers can only succeed when equal effort is placed on expanding the availability and quality of schooling (UNICEF Evaluation 80). At the level of schools and teachers, the evaluations described efforts to rapidly provide education during emergencies and reconstruct schools post-conflict (e.g., building temporary learning shelters, schools and teacher accommodation, the provision of textbooks and learning materials, recruitment of teachers, teacher training and coordination with partners including district government). Key reported challenges included the low quality of teaching, transitions into formal education, safety and security and lack of educational support for adolescents (UNICEF Evaluation 108). Evaluations found targeting approaches might not favour integration (refugee/host communities) and create friction between groups (WFP Evaluation 132). There is a need for a more upstream focus on building crisis preparation and risk-management into sector planning and the capacity of national governments. This will require further coordination, including across different international organisations. As crises continue, for example in Syria, evaluations report a need to switch from rapid response to resilience with longer-term planning needed (UNICEF Evaluation 79). Current challenges include struggles with transitioning to formal curriculum, restrictive enrolment policies and unclear pathways to adulthood for refugee and displaced youth.

8. Country-level responses to the study

110. To validate and enrich the findings of this evaluation synthesis, an early version of the study was presented at five country-level workshops held in Ghana, Guatemala, Lebanon, Nepal and Peru over the period from May 20 to June 11. Attendance was strong in all workshops with between 20 and 45 participants. Representation included national ministries of education, development partners and other national ministries and departments involved in monitoring and evaluation. A summary of the study findings, alongside an overview of evaluations in the dataset specific to their countries, was shared with all participants. Additional data were gathered through a short survey addressing the use, and potential for use, of programme evaluations in education in pursuit of SDG 4 Target 5.

111. National stakeholders reported that they primarily use domestic or national programme evaluations in their education policy-making and planning processes. They also cited the importance of national statistical data systems, where they often identified significant challenges and a need for ongoing support from development partners. In contrast, country stakeholders reported that they used evaluations from international organisations less frequently. In at least one country, stakeholders were unaware of the multiple international programme evaluations reported upon for their country in this evaluation synthesis. Nonetheless, workshop participants were confident that external programme evaluations could play a role of increasing importance in national policy debates and policy learning, with some modifications.

Box 8-1: Joint evaluation as best practice in Nepal

Nepal had among the largest number of evaluations of any country in this study (5), including several that reviewed the country's education sector programme. Yet these studies were not well-known by national stakeholders. Instead national stakeholders pointed to a joint evaluation of their sector programme not in our dataset as an example of best practice (Poyck, Koirala, Aryal and Sharma, 2016). This evaluation allowed the country to review their progress jointly with development partners. It was cited for its influence on national policies and the preparation of its new sector plan.

112. To this end, country stakeholders were eager to see how the evaluations sponsored by international organisations could be made more usable at the country-level to support their achievement of SDG 4 Target 5. They proposed the following areas for further improvement:

- ▶ More involvement of country stakeholders in the planning of external evaluations, including involvement in the identification of important questions, topics and in the design of studies;
- ▶ More coordination across international actors – with one country highlighting as a best practice a joint evaluation of their sector-wide programme;
- ▶ Creation of plain language summaries and other dissemination strategies to ensure wider uptake and use of evaluation evidence;
- ▶ Regular development of synthesised summaries of the evidence generated from both external and domestic evaluations to allow for better integration of different sources of evidence.

► Overall, there was good appetite for further collaboration on evaluation with international partners. Countries are eager to improve their ability to monitor and evaluate outcomes related to SDG 4 Target 5. More importantly, they are keenly aware of the equity challenges present in their national contexts and are motivated to learn from evidence on how to make progress in addressing them.

9. Conclusions and recommendations

113. **Since the adoption of the Sustainable Development Goals, significant work has been carried out to support improved data and indicators related to equality and inclusion in education.** The UNESCO Institute for Statistics (UIS), for example, has convened expert groups to work on indicator development, published methodological guidelines and expanded its international database with education indicators to improve coverage across countries, years and dimensions of disaggregation (UIS 2018a). UNESCO, GPE, UNICEF, the World Bank and other bilateral donors are also investing in national education management information systems that better capture data on equality and inclusion. Notable advances have occurred in specific areas, for example, new standards to support comparable statistics on disability sponsored by the Washington Group on Disability Statistics (Loeb, Mont, Cappa, Palma, Madans & Crialesi, 2018).
114. **However, the processes and mechanisms for monitoring and evaluating the equity impacts of different types of international educational interventions have evolved more slowly.** As noted in previous sections, during the MDG era, equality was often understood as equality in terms of access in donor-sponsored interventions, leading to an evaluative focus on counting numbers of schools and children, rather than on evaluating equality in terms of survival and learning. Common measures and standards for monitoring the equality in outcomes in education programmes and projects tended to cluster around a small number of indicators, like gender parity.
115. **The large volume and range of evaluations conducted with reference to educational equality and inclusion in different contexts has not been matched by a coordinated system to generate, synthesise and disseminate evaluation evidence related to equality and inclusion in education.** The evaluations in the dataset studied in this report are largely designed to provide accountability to the agencies who are responsible for implementing the activities or programmes or to donors who fund the specific activities under consideration. They are only indirectly designed as tools to support countries in their efforts to achieve SDG 4 Target 5. Nor are they explicitly structured to provide robust comparative evidence about what types of interventions will work to meet SDG 4 Target 5 or recommendations for replication and scale-up in different contexts.
116. Four key recommendations are proposed to help strengthen the effectiveness of the evaluations being conducted to meet these challenges.
117. **Address evidence gaps:** There is wide availability of evaluations of programme that target equity and inclusion; however, this is contrasted with uneven focus by level of intervention, target population, specific forms of exclusion and inequality and geography. Key gaps highlighted in the report include:
- Most evaluations are focused on interventions in primary and secondary education and early childhood education. They mainly report on equality in terms of access. They provide limited data on equality in vocational education, adult education and tertiary level education.
 - Evaluation evidence is strongest for interventions involving school feeding and conditional cash transfers. Evaluation evidence related to teachers and schools, including around system-wide school-based management reforms, is limited and relatively weak.
 - Too little evaluation evidence is available on equality and inclusion as related to children with disabilities, and those from indigenous, ethnic and linguistic minority backgrounds.
 - Geographically, the evaluations cover fewer than half of UN Member States; however, they provide data about education in sub-Saharan Africa and other regions often neglected in systematic reviews.

Recommendation 1 - Address evidence gaps: Agencies should work together to address areas where there are key gaps in evaluation evidence for SDG 4 Target 5. Such gap areas include: evaluations on teacher development, alternative/non-formal education as well as evaluations of the equity implications of school-based management, school grants and other decentralisation reforms. A particularly important imperative for future evaluations is the question of how complex system level interventions can be better designed to reach children and youth with disabilities, as well as indigenous and ethnic minority groups. Evaluation funding should also be focused on areas where there are hopeful qualitative but limited quantitative data to support scale-up of interventions.

118. **Contribute to stronger and more consistently available data:** Evaluations suffer from a lack of reliable, good quality, disaggregated data on those who are disadvantaged in education. Data on equality in terms of learning outcomes is especially lacking, as well as on disability and intersecting inequalities. Without such data, most programmes and interventions have trouble accurately identifying target populations; and evaluation cannot demonstrate that interventions are generating strong outcomes and impact. This problem cannot be solved on a once-off basis by collecting data within individual evaluations – it requires investments in national capacity and data-systems.

Recommendation 2 - Contribute to stronger and more consistently available data: Good quality national data on equality and inclusion is a global good. The evaluation offices of international organisations can support and contribute to national capacity by investing in the work of national partners when collecting data on equity, and supporting national stakeholders to engage in monitoring and formative evaluation. They can also join forces with national technical agencies and data units to ensure that national data systems are strengthened.

119. **Strengthen evaluation methodologies:** While the evaluations provide good qualitative evidence, the study highlights three main areas for improvement. First, there is limited use of robust experimental and quasi-experimental designs within evaluations conducted by the evaluation units of development agencies. Overall, evaluation findings related to cost-effectiveness and impact are also weakly reported in this dataset. At the same time, there is an urgent need to improve the uneven quality of evaluations of complex system-level interventions. This study highlights examples of good practice for evaluating complex, multi-component interventions in education, including the use of theories of change to explore causal links between different interventions, the use of systematic cross-national comparative methods and the potential to embed better data collection and impact evaluations within the monitoring and evaluation strategies at the time of programme design.

Recommendation 3 - Strengthen evaluation methodologies: Agencies should explore ways to strengthen their evaluations by i) incorporating stronger and more consistent measures of equality and inclusion; ii) embedding experimental evaluations into the design of system-level programmes; iii) creating stronger standards (OECD DAC, 2018; UNEG, 2014, 2016) for the evaluation of system-wide reform programme in education, including through the use of theory-based methodologies; and iv) invest in the collection of cost-effectiveness data. Creating a common set of 'best practices' for evaluation in these areas would lead to significant advances in the level of knowledge and evidence available related to SDG 4 Target 5.

120. **Synthesise and collaborate to make evidence more useful to national stakeholders:** The large volume and range of evaluations conducted with reference to equality and inclusion in education showcased in this report is not matched by a coordinated system to generate, synthesise and disseminate evaluation evidence in a manner that allows for aggregation and learning. The evaluations in this study are not primarily designed to build a shared knowledge base about what works for achieving SDG 4 Target 5 across countries and organisations. Instead, evaluation questions are focused on understanding the organisational effectiveness of the agencies responsible for delivering the intervention or programme, with the sponsoring organisation as their main audience. Few evaluations are conducted jointly.

Recommendation 4 - Synthesise and collaborate to make evidence more useful to national stakeholders:

Building on the momentum generated through this collaborative study, and on recent UN commitments for greater collaboration and coordination of evaluation work (UNGA, 2019), development partners should consider ways in which they can systematically coordinate, synthesise and promote learning from their evaluations. Effective strategies might include the creation of a registry of education evaluations planned or commissioned; commitment to regular international synthesis of evaluations; and routine efforts to link evaluation evidence to the wider research literature on what works. Even more importantly, evaluation offices should consider ways in which they can work more closely with governments and other national stakeholders to identify key areas where there are gaps in evaluation and to collaborate on all aspects of evaluation from design and identification of questions to learning and dissemination.

121. The overall purpose of this report was to understand how education evaluations undertaken by UN organisations can support the achievement of SDG 4 Target 5. As the report has shown, evaluations funded by UN and other development agencies already play a critical role in ensuring international organisations are accountable to their donors and stakeholders. They present real-world examples of educational initiatives and lessons learned from them and may inform policy-makers and other decision-makers at global, national and local levels.
122. International organisations now need to take important new steps to ensure that evaluations are more explicitly designed to contribute to a common evidence-base on what works in achieving equality and inclusion in education and why. Better engagement of national stakeholders in evaluation design and more coordination among global actors can help to ensure learning and progress on equality and inclusion in education. Organisations should consider further investments in coordination, synthesis, national engagement and learning. This study provides a starting point for such future collaboration.

Reference list

3ie (International Initiative for Impact Evaluation). 2010. Conditional cash transfer programmes: A magic bullet to improve people's health and education? (Impact Evaluation 01).

3ie. 2019. Primary and Secondary Evidence Gap Map. Retrieved from: <http://gapmaps.3ieimpact.org/evidence-maps/primary-and-secondary-education-evidence-gap-map>

Government of Australia. 2017. What works best for Education for Development: A Super Synthesis of the Evidence. Barton, Australia: DFAT. Retrieved from: <https://dfat.gov.au/about-us/publications/Documents/super-synthesis-report.pdf>

Aurino, E., Tranchant, J.-P., Sekou Diallo, A. and Gelli, A. 2018. School Feeding or General Food Distribution? Quasi-Experimental Evidence on the Educational Impacts of Emergency Food Assistance during Conflict in Mali (Innocenti Working Paper No. No. WP-2018-04). UNICEF Office of Research. Retrieved from: <https://www.unicef-irc.org/publications/pdf/Aurino%20et%20al%2>

Bashir, S, Lockheed, M.; Ninan, E. and Tan, J.P. 2018. Facing Forward: Schooling for Learning in Africa. Africa Development Forum series. Washington, DC: World Bank. Retrieved from: <https://openknowledge.worldbank.org/handle/10986/29377>

Bauchet, J., Undurraga, E. A., Reyes-García, V., Behrman, J. R. and Godoy, R. A. 2018. Conditional cash transfers for primary education: Which children are left out? World Development, 105, 1–12. Retrieved from: <https://doi.org/10.1016/j.worlddev.2017.12.021>.

Bundy, A. P., de Silva, N, Horton, S. and Patton G.C.. 2018. Investment in child and adolescent health and development: Key messages from Disease Control Priorities, 3rd edition. The Lancet, 391(10121), 687-699. Retrieved from: [https://doi.org/10.1016/S0140-6736\(17\)32417-0](https://doi.org/10.1016/S0140-6736(17)32417-0)

Burde, D., Guven, O., Kelcey, J., Lahmann, H. and Al-Abbadi, K. 2015. What Works to Promote Children's Educational Access, Quality of Learning, and Wellbeing in Crisis-Affected Contexts. Education Rigorous Literature Review. London, UK: Department for International Development. Retrieved from: <http://s3.amazonaws.com/inee-assets/resources/Education-emergencies-rigorous-review-2015-10.pdf>

Carr-Hill, R., Rolleston, C., Pherali, T. and Schendel, R. 2014. The Effects of School-Based Decision Making on Educational Outcomes in Low and Middle Income Contexts: A Systematic Review. London, UK: Department for International Development. Retrieved from: <https://campbellcollaboration.org/library/school-based-decision-making-in-low-and-middle-income-contexts.html>

Clist, P. 2018. Payment by results in international development: Evidence from the first decade. Development Policy Review. Accepted Author Manuscript. doi:10.1111/dpr.12405.

Dahlgren, S. 2007. Evaluations of Country Strategies: An Overview of Experiences and a Proposal for Shaping Future of Country Programme Evaluations. SIDA Studies in Evaluation 2007:4. Stockholm: SIDA. Retrieved from: <https://www.oecd.org/derec/sweden/38971661.pdf>

Donnelly, C.A., Boyd, I., Campbell, P., Craig, C., Vallance, P., Whitty, C.J.M., et al. 2018. Four principles for synthesising evidence. Nature 558(7710), 361–4.

Dryden-Peterson, S. 2018. Inclusion of refugees in national education systems. Background Paper for the Global Education Monitoring Report. GEMR/MRT/2018/P1/6.

Eddy-Spicer, D., Ehren, M., Mukdarut, B., Khatwa, M. and Perrone, F. 2016. Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low- and middle-income countries: Systematic review. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. Retrieved from: <http://eppi.ioe.ac.uk/CMS/Portals/0/PDF%20reviews%20and%20summaries/Q4%20System%20efficiency.pdf>

Education Equity Research Initiative. 2017. Measuring Equity in Education: Review of the Global and Programmatic Landscape. Washington, DC: EERI.

Education Equity Research Initiative. 2017. Practical Recommendation for Equity Analysis in Education. Washington DC: EERI. Retrieved from: <http://www.educationequity2030.org/resources-2/2017/3/8/practical-recommendations-for-equity-analysis-in-education>

Evans, D., and Popova, A. 2016. What Really Works to Improve Learning in Developing Countries?: An Analysis of Divergent Findings in Systematic Reviews. World Bank Research Observer, 31(2), 242–270.

Evans, D. and Yuan, F. 2019. We Can Learn a Lot about Improving Girls' Education from Interventions That Don't Target Girls. Washington, D.C.: Center for Global Development. Retrieved from: <https://www.cgdev.org/blog/we-can-learn-lot-about-improving-girls-education-interventions-dont-target-girls>.

Gelli, A., Masset, E., Folsom, G., Kusi, A., Arhinful, D. K., Asante, F., Drake, L. et al. (2016). Evaluation of alternative school feeding models on nutrition, education, agriculture and other social outcomes in Ghana: rationale, randomised design and baseline data. *Trials*, 17, 37. doi:10.1186/s13063-015-1116-0.

ICAI (Independent Commission for Aid Impact). 2018. ICAI's Follow-Up of: Accessing, Staying, Learning – UK Aids Support to Marginalized Girls. Retrieved from: <https://icai.independent.gov.uk/wp-content/uploads/ICAI-Follow-up-Marginalised-Girls.pdf>

IOD Parc. 2012. Meta Evaluation of Mine Action and Development. Sheffield, UK: IOD Parc. Retrieved from : <https://www.iodparc.com/documents/metaevaluation.pdf>

J-PAL. 2017. Roll Call: Getting Children Into School (J-PAL Policy Bulletin). Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL). Retrieved from: <https://www.povertyactionlab.org/sites/default/files/publications/roll-call-getting-children-into-school.pdf>

Mayne, J. 2008. Contribution Analysis: An approach to exploring cause and effect, ILAC Methodological. Retrieved from: http://www.cgiarilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. and PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6:e1000097.

Langlois, E.V., Daniels, K. and Akl, E.A. (2018). Evidence synthesis for health policy and systems: a methods guide. Geneva: World Health Organization.

Loeb, M, Mont, D., Cappa, C., Palma E.D, Madans J. and Cialesi, R. 2018. The development and testing of a module on child functioning for identifying children with disabilities on surveys. *Disability and Health Journal* 11, 495-501.

Magrath, B., Aslam, M. and Johnson, D. 2019. Systems research in education: Designs and methods. *Research in Comparative and International Education*, 14(1), 7-29.

OECD DAC. 1991. Principles for Evaluation of Development Assistance. Paris: OECD. Retrieved from: <https://www.oecd.org/dac/evaluation/dcdndep/45438179.pdf>

OECD DAC. 2018. OECD DAC Evaluation Criteria: Summary of consultation responses. Paris, France: OECD. Retrieved from: <https://ieg.worldbankgroup.org/sites/default/files/Data/DAC-Criteria/ConsultationReportEvaluationCriteria.pdf>

Popova, A., Evans, D., K., Breeding, M., E. and Arancibia, V. 2018. Global Landscape of In Service Teacher Professional Development Programs: The Gap between Evidence and Practice (Policy Research Working Paper No. 8572). Washington, DC: RISE Research, Centre for Global Development. Retrieved from: <https://openknowledge.worldbank.org/bitstream/handle/10986/30324/WPS8572.pdf?sequence=1&isAllowed=y>

Poyck, C., Koirala, B.N., Aryal, P.N, and Sharma, N.K. 2016. Joint Evaluation of Nepals School Sector Reform Program, 2009-2016. Brussels: European Union.

Rose, P. and Alcott, B. 2015. How Can Education Systems Become More Equitable? DFID Think Piece. London: HEART. Retrieved from: <http://uis.unesco.org/sites/default/files/documents/how-can-education-systems-become-equitable-by-2030-learning-and-equity-pauline-rose-benjamin-alcott-heart-2015-en.pdf>

Rose, P. and Yorke, L. 2019. General versus Girl-Targeted Interventions: A False Dichotomy? A response to Evans and Yuan. REAL Centre, University of Cambridge. Retrieved from: <https://doi.org/10.5281/zenodo.3249882>

Samoff, J., Leer, J. and Reddy, M. 2016. Capturing Complexity and Context: Evaluating Aid to Education. Stockholm, Sweden: Expertgruppen för biståndsanalys (EBA).

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M. and Jimenez, E. 2016. The impact of education programmes on learning and school participation in low- and middle-income countries: a systematic review summary report (3ie Systematic Review Summary No. 7). London: International Initiative for Impact Evaluation (3ie). Retrieved from: <http://www.3ieimpact.org/evidence-hub/publications/systematic-review-summaries/impact-education-programmes-learning-school-participation-low-and-middle-income-countries>

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T. and Evers, J. 2015. Interventions for improving learning outcomes and access to education in low-and middle-income countries: a systematic review. Retrieved from: <http://www.3ieimpact.org>

Tranchant, J.-P., Gelli, A., Bliznashkab, L., Sekou Diallo, A., Sacko, M., Assima, A. and Masset, E. 2019. The impact of food assistance on food insecure populations during conflict: Evidence from a quasi-experiment in Mali. World Development, 119, 185–202. <https://doi.org/10.1016/j.worlddev.2018.01.027>

Tricco, A., Langlois, E.V. and Strauss, S. 2017. Rapid Reviews to Strengthen Health Policy and Systems: A practical guide. Geneva: World Health Organisation.

UNEG (United Nations Evaluation Group). 2014. Integrating Human Rights and Gender Equality in Evaluations. New York: UNEG. Retrieved from: <http://www.unevaluation.org/document/download/2107>

UNEG (United Nations Evaluation Group). 2016. Norms and Standards for Evaluation. New York, NY: UNEG. Retrieved from: <http://www.unevaluation.org/document/detail/1914>

UNESCO. 2017. UNESCO's Education Evaluations in 2016: A Review. Paris: UNESCO.

UNESCO GEM. 2018. Global Education Monitoring Report). Migration, Displacement and Education. Paris: UNESCO.

UNESCO GMR. 2015. Education For All, 2000-2015 – Achievements and Challenges. Paris: UNESCO.

UIS (UNESCO Institute for Statistics). 2018a. Quick Guide to the Education Indicators for SDG 4. Montreal, Québec. Retrieved from: <http://uis.unesco.org/sites/default/files/documents/quickguideeducationindicators-sdg4-2018-en.pdf>

UIS. (UNESCO Institute for Statistics). 2018b. Handbook on Measuring Equity in Education. Paris, France: UNESCO. Retrieved from: <http://uis.unesco.org/sites/default/files/documents/handbook-measuring-equity-education-2018-en.pdf>

UIS. (UNESCO Institute for Statistics). 2018c. SDG 4 Data Digest 2018: Data to Nurture Learning. Montreal, Québec: UNESCO-UIS. Retrieved from: <http://uis.unesco.org/sites/default/files/documents/sdg4-data-digest-data-nurture-learning-2018-en.pdf>

UNGA (United Nations General Assembly). 2019. Implementation of the Gender Assembly Resolution 71/243 on the quadrennial review of the operational activities for the development of the United Nations System. 2019: Funding Compact. 74th Session, Session Document A/74/73/Add.1: Retrieved from: <https://www.un.org/ecosoc/en/node/17356550>

USAID (United States Agency for International Development). 2018. Synthesis of Findings and Lessons Learned from USAID-Funded Evaluations: Education Sector, 2013-2016. Washington, D.C.: USAID.

USAID. 2017. Tool: Assessing the Quality of Evaluations. Office of Education. Washington D.C: USAID.

Weiss, C. H. 1997. Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 76, 41–55.

White, H. 2009. Theory-based impact evaluation: Principles and Practice (Working Paper 3). International Initiative for Impact Evaluation (3ie).

Appendix A: List of included evaluations

(Listed by organisation and number in the total set of 147 organizations)

DFID

1. Cambridge Education. (2015). Evaluation: Pilot Project of Results-Based Aid in the Education Sector, Ethiopia. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/576168/Eval-pilot-project-result-based-aid-Education-Ethiopia.pdf
2. Coffey. (2017a). Endline Evaluation Report Strategic Partnerships Window. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/701007/Endline-report-Girls-Education-Challenge-Partnership-Window.pdf
3. Coffey. (2017b). Girls Education Challenge Fund Innovation Window: Endline report. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/700823/Girls-Education-Challenge-Fund-Innovation-Window-Endline-Report.pdf
4. Coffey. (2017c). Step Change Window Endline Evaluation Report. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/676532/GEC-EM-SCW-Endline-Evaluation-Report-final.pdf
5. Ecorys. & Fiscus. (2016). Evaluation of Budget Support to Sierra Leone 2002 to 2015: volume 1. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/563181/Budget-support-Sierra-Leone-volume1.pdf
6. Montrose. (2016). Evaluation of Youth Development and Northern Uganda Youth Entrepreneurship Programme. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/607549/Evaluation-of-Youth-Development-and-Northern-Uganda-Youth-Entrepreneurship-Programme.pdf
7. Mott MacDonald. (2017). Evaluation of Adolescent Girls Empowerment Programme - Zambia. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/607393/Evaluation-of-Adolescent-Girls-Empowerment-Programme.pdf
8. Upper Quartile. (2015). Evaluation of Results Based Aid in Rwandan Education: Final Evaluation Report. DFID. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/563272/Results-Based-Aid-Rwandan-Education-final-eval.pdf

GLOBAL PARTNERSHIP FOR EDUCATION

9. Oxford Policy Management. (2018). Evaluation of the Global Partnership for Education (GPE)' s Support for Civil Society Engagement - Final Report. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/evaluation-global-partnership-educations-support-civil-society-engagement-final-report>
10. Universal Management Group. (2018a). Summative evaluation of GPE's country-level support: Cote d'Ivoire. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-cote-divoire>
11. Universal Management Group. (2018b). Summative evaluation of GPE's country-level support to education: Burkina Faso. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-education-burkina-faso>
12. Universal Management Group. (2018c). Summative evaluation of GPE's country-level support to education: Guyana. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-education-guyana>

evaluation-gpes-country-level-support-education-guyana

13. Universal Management Group. (2018d). Summative evaluation of GPE's country-level support to education: Liberia. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-education-liberia>

14. Universal Management Group. (2018e). Summative evaluation of GPE's country-level support to education: Pakistan. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-education-pakistan>

15. Universal Management Group. (2018f). Summative evaluation of GPE's country-level support to education: Sierra Leone. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-education-sierra-leone>

16. Universal Management Group. (2018g). Summative Evaluation of GPE's Country-level Support to Education: The Gambia. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-country-level-support-education-gambia>

17. Universal Management Group. (2018h). Summative evaluation of GPE's country-level support to education: Mauritania. Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/summative-evaluation-gpes-support-mauritania>

18. Universal Management Group. (2018i). GPE Country-level Evaluations: Synthesis Report (No. Global Partnership for Education). Global Partnership for Education. Retrieved from <https://www.globalpartnership.org/content/synthesis-report-gpe-country-level-evaluations-february-2019>

IED - ASIAN DEVELOPMENT BANK

19. IED ADB. (2016). Sri Lanka: Distance Education Modernization Project. Asian Development Bank. Retrieved from <https://www.adb.org/documents/sri-lanka-distance-education-modernization-project-1>

20. Independent Evaluation. (2017). Asian Development Bank Support for Gender and Development (2005–2015). Asian Development Bank. Retrieved from <https://www.adb.org/sites/default/files/evaluation-document/181135/files/tes-gender-and-development.pdf>

21. Independent Evaluation Department. (2016). Country Assistance Programme Evaluation for Sri Lanka (2006-2015). Asian Development Bank. Retrieved from <https://www.adb.org/sites/default/files/evaluation-document/175387/files/cape-sri.pdf>

INDEPENDENT EVALUATION GROUP – WORLD BANK

22. Independent Evaluation Group. (2015). World Bank Support to Early Childhood and Development. International Bank for Reconstruction and Development/World Bank. Retrieved from https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/early_child_dev_eval.pdf

23. Independent Evaluation Group. (2015). Project Performance Assessment Report: Nicaragua Education Project (IDA-39780; TF-53991; TF-57311). World Bank. Retrieved from <http://documents.worldbank.org/curated/en/480621468179086433/pdf/97131-PPAR-P078990-P150246-SecM2015-0220-Box391496B-OUO-9.pdf>

24. Independent Evaluation Group. (2017a). Brazil: Pernambuco Education Results and Accountability Project. International Bank for Reconstruction and Development/World Bank. Retrieved from <https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ppar-brazil-07142017.pdf>

25. Independent Evaluation Group. (2017b). Higher Education for Development: An Evaluation of the World Bank Group's Support. International Bank for Reconstruction and Development/World Bank. Retrieved from <https://openknowledge.worldbank.org/bitstream/handle/10986/26486/113867-WP-PUBLIC.pdf?sequence=5&isAllowed=y>

26. Independent Evaluation Group. (2017c). Kyrgyz Republic: Rural Education Project. International Bank for Reconstruction and Development/World Bank. Retrieved from <http://ieg.worldbankgroup.org/sites/default/files/Data/reports/ppar-kyrgyzeducation-07172017.pdf>
27. Independent Evaluation Group. (2015). Project Performance Assessment Report: Nepal Education for All (IDA-39560 IDA-H3400). World Bank Group. Retrieved from <http://documents.worldbank.org/curated/en/258151468196129867/pdf/101692-PPAR-P074633-P150605-IDA-SecM2016-0010-OUO-9.pdf>
28. Independent Evaluation Group. (2015). Indonesia - Managing Higher Education for Relevance and Efficiency Project (English). World Bank. Retrieved from <http://documents.worldbank.org/curated/en/185401468188648046/pdf/97132-PPAR-P085374-SecM2015-0221-Box391488B-OUO-9.pdf>
29. Independent Evaluation Group. (2016). Malawi - Third Social Action Fund and Second Phase of Adaptable Programme Lending Project (English). International Bank for Reconstruction and Development/World Bank. Retrieved from <http://documents.worldbank.org/curated/en/537571492613306986/pdf/113148-PPAR-P078113-REPLACEMENT-PUBLIC.pdf>
30. Independent Evaluation Group. (2017). Project Performance Assessment Report: Lao People's Democratic Republic Second Education Development Project. International Bank for Reconstruction and Development/World Bank. Retrieved from <http://documents.worldbank.org/curated/en/537571492613306986/pdf/113148-PPAR-P078113-REPLACEMENT-PUBLIC.pdf>
31. Independent Evaluation Group. (2018b). Project Performance Assessment Report: Bulgaria Social Inclusion Project. International Bank for Reconstruction and Development/World Bank. Retrieved from https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ppar_bulgariasocialinclusion.pdf
32. Independent Evaluation Group. (2015). Vietnam - Sixth-Tenth Poverty Reduction Supports Credit Project (Project Performance Assessment Report). World Bank. Retrieved from <http://documents.worldbank.org/curated/en/721331467986245693/pdf/96203-PPAR-REVISED-PUBLIC-VIETNAM-PPAR-FINAL.pdf>
33. Independent Evaluation Group. (2018). Project Performance Assessment Report: Philippines Support for Basic Education Reform, 2006-12. International Bank for Reconstruction and Development/World Bank. Retrieved from https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ppar_philippineseducationreform.pdf

INTERAMERICAN DEVELOPMENT BANK

34. Office of Evaluation and Oversight (OVE). (2018). Evaluation of the Bank's Support for Gender and Diversity. InterAmerican Development Bank. Retrieved from <https://publications.iadb.org/en/publication/17380/evaluation-banks-support-gender-and-diversity>

NORAD

35. Development Portfolio Management Group. (2015a). Evaluation of Norwegian Multilateral Support to Basic Education. Norad. Retrieved from <https://www.oecd.org/derec/norway/Evaluation-Norwegian-Multilateral-Support-Basic-Education.pdf>
36. Development Portfolio Management Group. (2015b). Evaluation of Norwegian Multilateral Support to Basic Education: Ethiopia Case Study. Norad. Retrieved from <https://norad.no/globalassets/publikasjoner/publikasjoner-2015-/evaluering/evaluation-of-norwegian-multilateral-support-to-basic-education/7.15-ethiopia-case-study.pdf>
37. Development Portfolio Management Group. (2015d). Evaluation of Norwegian Multilateral Support to Basic Education: Madagascar Case Study. Retrieved from <https://norad.no/globalassets/publikasjoner/publikasjoner-2015-/evaluering/evaluation-of-norwegian-multilateral-support-to-basic-education/7.15-madagascar-case-study.pdf>
38. Development Portfolio Management Group. (2015f). Evaluation of Norwegian Multilateral Support to Basic Education: Malawi Case Study. Norad. Retrieved from <https://norad.no/globalassets/publikasjoner/>

publikasjoner-2015-/evaluating/evaluation-of-norwegian-multilateral-support-to-basic-education/7.15-malawi-case-study.pdf

39. Development Portfolio Management Group. (2015h). Evaluation of Norwegian Multilateral Support to Basic Education: Nepal Case Study. Retrieved from <https://norad.no/globalassets/publikasjoner/publikasjoner-2015-/evaluating/evaluation-of-norwegian-multilateral-support-to-basic-education/7.15-nepal-case-study.pdf>

40. Betts, J., Ghanem, N., Al Baba, D., and Azar, R. (2017). Evaluation of Norway's Support to Education in Conflict and Crisis through Civil Society Organisations: Lebanon Case Study. Norad. Retrieved from https://norad.no/contentassets/5490809999e545be8257166aaa01ffd9/lebanon_case-study-report.pdf

41. Hamar, M.C., Abudoha, R., Omer, Y., and Gubser, M. (2017). Evaluation of Norway's Support to Education in Conflict and Crisis through Civil Society Organisations: Jordan Case Study Report. Norad. Retrieved from https://norad.no/contentassets/5490809999e545be8257166aaa01ffd9/jordan_case-study-report.pdf

42. O'Grady, N. (2017). Evaluation of Norway's Support to Education in Conflict and Crisis through Civil Society Organisations: South Sudan Desk Study Report. Norad. Retrieved from https://norad.no/contentassets/5490809999e545be8257166aaa01ffd9/south-sudan_desk-study-report.pdf

43. Olsen, K. (2017). Evaluation of Norway's Support to Education in Conflict and Crisis through Civil Society Organisations: Somalia Desk Study Report. Norad. Retrieved from https://norad.no/contentassets/5490809999e545be8257166aaa01ffd9/somalia_desk-study-report.pdf

44. The Konterra Group. (2017). Realising Potential: Evaluation of Norway's Support to Education in Conflict and Crisis through Civil Society Organisations. Norad. Retrieved from https://norad.no/contentassets/5490809999e545be8257166aaa01ffd9/9.17-education-in-conflict-and-crisis_evaluation-report.pdf

UNESCO

45. Broek, S., Werquin, P., Coles, M., and Buiskool, B.J. (2015). Evaluation of the UNESCO Thematic Area "TVET" (Technical and Vocational Education and Training). UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000234442.locale=en>

46. Colman, R., and Ngoc Nga, D. (2018). Gender Equality and Girls' Education Initiative in Viet Nam: Empowering girls and women for a more equal society. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000367489>

47. Dietrich, C., Hénard, F., and Werquin, P. (2018). Évaluation de l'IIEP-Pôle de Dakar. UNESCO. Retrieved from http://www.iiep.unesco.org/sites/default/files/pdk_2018_eval_rapport_final_forwebsite.pdf

48. Ipsos MORI. (2018). Evaluation of the Global Education Monitoring Report: Final Report v3. UNESCO. Retrieved from https://en.unesco.org/gem-report/sites/gem-report/files/External_Evaluation.pdf

49. Klein, T. (2017). End-of-Project Evaluation of the Korea Government-funded Project: Technical & Vocational Education & Training for Syrian & Jordanian Youth in Jordan. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000367450>

50. MDF Training and Consultancy. (2017). External Evaluation Technical Cooperation Programme IIEP Paris. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000366601>

51. Samaran, V., and Dilawar, E. (2018). Afghanistan: Capacity Development for Educational Planning and Management III (CD-EPM-III): April 2015 – March 2018. UNESCO. Retrieved from http://www.iiep.unesco.org/sites/default/files/iiep-unesco_evaluation_afghanistan_and_management_response.pdf

52. UNESCO Internal Oversight Service. (2017). Evaluation of UNESCO's programme interventions on girls' and women's education. Paris, France: UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000258978?posInSet=1&queryId=N-930b8bc9-f400-4557-b3cd-5789abf06fe4>

53. UNESCO Internal Oversight Service. (2016a). Evaluation of UNESCO's Role in Education in Emergencies and

Protracted Crises: Case Study 2 - Crisis-Sensitive Education Sector Planning: UNESCO-IIEP Support in South Sudan. UNESCO. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/IOS/images/Evaluation_EiE_CS2_SSsudan.pdf

54. UNESCO Internal Oversight Service. (2016b). Evaluation of the Education for All (EFA) Global and Regional Coordination Mechanisms. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000245299>

55. UNESCO Internal Oversight Service (2016c). Evaluation of UNESCO's Role in Education in Emergencies and Protracted Crises. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000246095>

56. UNESCO Internal Oversight Service. (2016e). Evaluation of UNESCO's Role in Education in Emergencies and Protracted Crises: UNESCO Kathmandu Office's Education Response to Natural Disaster in Nepal. UNESCO. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/IOS/images/Evaluation_EiE_CS3_Nepal.pdf

57. UNESCO Internal Oversight Service. (2016). Evaluation of UNESCO's Role in Education in Emergencies and Protracted Crises: UNESCO's Education Response to the Syria Crisis: Towards Bridging the Humanitarian-Development Divide [Case Study 1]. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000246279>

UNDP

58. Centre for Continuing Education, Birzeit University. (2018). Right to Education in East Jerusalem: Final Project Evaluation Report. UNDP. Retrieved from <https://erc.undp.org/evaluation/evaluations/detail/8944>

59. Jamaspishvili, T. (2017). Deepening linkages between formal/non-formal VET system and the labour market needs in the context of lifelong learning in Georgia. UNDP. Retrieved from <https://erc.undp.org/evaluation/documents/detail/11439>

60. Sabina, N., and Chakma, K.N. (2015). Promotion of Development and Confidence Building in the Chittagong Hill Tracts. UNDP. Retrieved from <https://erc.undp.org/evaluation/evaluations/detail/7971>

UNHCR

61. Ambroso, G., Meier-Ewert, G.S., Parker, J. and Richardson, L. (2016). Evaluation of UNHCR's Response to the L3 South Sudan Refugee Crisis in Uganda and Ethiopia. UNHCR. Retrieved from <https://www.unhcr.org/research/evalreports/56b1d8df9/evaluation-unhcrs-response-l3-south-sudan-refugee-crisis-uganda-ethiopia.html>

62. Oxford Policy Management. (2017). Evaluation of UNHCR's implementation of three of its protection strategies: The Global Education Strategy, the Updated SGBV Strategy, and the Child Protection Framework. UNHCR. Retrieved from <https://www.unhcr.org/5a183d9c7.pdf>

63. Universalia. (2016). Evaluation of UNHCR's Emergency Response to the influx of Syrian Refugees into Turkey. UNHCR. Retrieved from <https://www.unhcr.org/research/evalreports/58a6bbca7/evaluation-unhcrs-emergency-response-influx-syrian-refugees-turkey-full.html>

UNICEF

64. UNICEF (2018). Philippines: Evaluation of UNICEF Tahderiyah Programme (2010-2017) funded by DFAT. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/TECD_Evaluation_Report_Final_Philippines_2018-001.pdf

65. UNICEF. (2016). Malawi Social Cash Transfer Programme Endline Impact Evaluation Report. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Malawi_Social_Cash_Transfer_Programme_Endline_Impact_Evaluation_Report_Malawi_2016-005.pdf

66. UNICEF (2017). Evaluation Report of the UNICEF Ghana Education Programme (2012-2017): A Capacity Building Perspective. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/AIR_UNICEF_Ghana_CD_

Report_revised_final.pdf

67. UNICEF (2016). Evaluación de impacto de la estrategia de gestión educativa integral para el ámbito municipal. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_92854.html
68. UNICEF (2017). The Evaluation report of Meshwary Project Phase II. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Meshwary_Project_Phase_II_-_Final_Evaluation_Report_-_ECO.pdf
69. UNICEF. (2015). Evaluation of the Social Cash Transfer Pilot Programme, Tigray Region, Ethiopia: Endline Report. Ethiopia: UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/UNICEF_Tigray_Endline_Report_FINAL_Ethiopia_2015-051.pdf
70. UNICEF. (2015). Madagascar Evaluation du Contrat Programme de Reussite Scolaire CPRS. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Unicef_Madagascar_Rapport_evaluation_CPRS_final_2015-002.pdf
71. UNICEF. (2017). Evaluation of the UNICEF Philippine Country Office “Early Childhood Care and Development” and “Basic Education” components of the 7th GPH-UNICEF Country Programme 2012-2016. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_98462.html
72. UNICEF. (2018). Evaluation of Global Partnership for Education Programme in South Sudan. UNICEF. Retrieved from <https://www.unicef.org/evaldatabase/files/SouthSudan-FinalReport-2018-001.pdf>
73. UNICEF. (2017). Impact Evaluation of UNICEF Nigeria Girls’ Education Project Phase 3 (GEP3) Cash Transfer Programme (CTP) in Niger and Sokoto States. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/NGA_GEP3_Impact_Evaluation_Report_final_11-12-2017.pdf
74. UNICEF. (2017). An Impact Evaluation of Alternative Basic Education in Ethiopia: Final Evaluation Report. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_101831.html
75. UNICEF. (2015). Burkina Faso External Evaluation of Quality Child-friendly Schools in the province of Ganzourgou and Namentenga, Burkina Faso. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/BurkinaFaso_2015-009_Rapport_EQAmE_FINAL_FINAL.pdf
76. UNICEF. (2017). Evaluation de la mise en œuvre des activités génératrices de revenus par les associations de mères éducatrices. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_100955.html
77. UNICEF. (2016). Sri Lanka Child-Friendly Approach (CFA) Evaluation. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_90725.html
78. UNICEF. (2018). Evaluation of UNICEF Girls’ Education Portfolio (2009-2015). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_103356.html
79. UNICEF. (2015). Evaluation of UNICEF’s Emergency Education Response for Syrian Refugee Children and Host Communities in Jordan. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/UNICEF_Jordan_Education_response_Evaluation_2015_Final_report.pdf
80. UNICEF. (2018). “Min Illa” Cash Transfer Programme for Displaced Syrian Children in Lebanon. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/2018-001-Lebanon-UNICEF_Impact_Evaluation_Report_Endline.pdf
81. UNICEF. (2015). Evaluation of Meena Radio Programme: India. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Evaluation_of_Meena_Radio_Programme_Final_Report_India_2015-002.pdf
82. UNICEF. (2015). Evaluation of Techno Girl: A Mentoring and Job-Shadowing Programme for Girls Studying Science, Technology, Engineering and Maths Subjects in South Africa. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Technogirls_final_2015-001_South_Africa.pdf

83. UNICEF. (2015). Evaluation of Activity-Based Learning as a means of Child-Friendly Education. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/EI-UNICEF_Evaluation_of_ABL_as_a_means_of_child-friendly_education-Final_Report_India_2015-024.pdf
84. UNICEF (2016). Endline Evaluation: UNICEF Kenya's Implementation of Integrated Interventions Under the Child Friendly Schools Framework in ASAL Counties. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/SHWETA_UNICEF_ASAL_IMPACT_EVALUATION_FINAL_REPORT_Kenya_2016-007.pdf
85. UNICEF. C. (2016). Evaluation of UNICEF Tajikistan's Work in Priority Districts During the 2010-2015 Country Programme. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Tajikistan_FinalEvalReport_vol1_2015-001.pdf
86. UNICEF. (2017). Evaluation of the Programme "Montenegro – Investment case on Early Childhood Development": Final Evaluation Report (Vol. 1). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/MNE_ECE_Investment_Case_FinalEvalReport_vol1_Montenegro_2017-001.pdf
87. UNICEF. (2016). Evaluation of the Youth Education Pack (YEP) Programme in Somalia. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/YEP_Evaluation_in_Somalia_-_Final_Report_2016-002.pdf
88. UNICEF. (2016). Evaluation of the "I believe in you" In-Service Teacher Training Programme in the Interior of Suriname. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_92859.html
89. UNICEF. (2016). Evaluation of Community Preschool Modality in Cambodia April-December 2015 Final Report (Volume 1). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Final_Evaluation_Report_CPS_Modality-Volume_I_Cambodia_2016-001.pdf
90. UNICEF. (2015). Rapport final d'évaluation du Programme "Appui d'Urgence à l'Éducation Nationale Malgache (AUENM)" UNICEF et PAM. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Rapport_final_eval_AUENM_Madagascar_2015-001.pdf
91. UNICEF. (2015). Evaluation of Child Friendly Schools (CFS) Teacher Training Intervention Under UNICEF-Supported CFS Initiative of Ministry of Education in Timor-Leste. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/UNICEF_Timor-Leste_Evaluation_of_Child_Friendly_Schools_-_FINAL_2015-002.pdf
92. UNICEF. (2016). Communication for Development: An Evaluation of UNICEF's Capacity and Action: Viet Nam Country Case Study. Vietnam. Retrieved from https://www.unicef.org/evaldatabase/index_95490.html
93. UNICEF. (2018). Evaluation of the National Early Childhood Development Programme: Final Report. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Evaluation_of_the_National_ECD_Programme_OPM_full_report_final_Nepal_2018-001.pdf
94. UNICEF. (2016a). Evaluation of the GEMA Programme School Management for Learning Improvement. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_92858.html
95. UNICEF. (2016b). Evaluation of the Programme ICT mediated secondary rural schools. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_92857.html
96. UNICEF. (2018). Summative Evaluation Report: "Investing in Education and Building Resilience in Gaza." UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_103289.html
97. UNICEF. (2016). Evaluación de los Programas de Premedia Multigrado y Telebásica en Panamá. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_95493.html
98. UNICEF. (2015). Evaluation of the Conditional Cash Transfer Programme Avancemos. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Informe_Final_Evaluacion_AVANCEMOS_CostaRica_2015-001.pdf
99. UNICEF. (2015). United Republic of Tanzania/UNICEF Evaluation of TUSEME Programme in Tanzania.

UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_90229.html

100. UNICEF. (2015). Evaluation of the UNICEF Supported Adult Literacy Programme in Guinea Bissau. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_92871.html

101. UNICEF. (2015). CAR Evaluation du Projet de Soutien au redemarrage et au suivi des activites educatives en Republique Centrafricaine. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Rapport_Final_Evaluation_du_GPE_CAR-10-2015.pdf

102. UNICEF. (2018). Vietnam: Evaluation of the Project on “Strengthening Capacity for Child-sensitive Planning, Monitoring and Evaluation for Socio-economic Development Plans.” UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Viet_Nam_Evaluation_Report_on_SEDP_Project_FINAL_2018-002.pdf

103. UNICEF. (2015). Evaluation of the Equity Focus of the Malaysia Country Programme. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Evaluation_Report_-_Design_-_Consolidated_15_Dec_Update_Malaysia_2015-001.pdf

104. UNICEF. (2018). Tanzania Youth Study of the Productive Social Safety Net (PSSN) Impact Evaluation: Endline Report. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/PSSN_Youth_Endline_Report_-_April_2018_Tanzania_2018-001.pdf

105. UNICEF. (2017). Evaluation of the “Breaking the cycle of exclusion for Roma children through Early Childhood Development and Education” multi-country project in the Former Yugoslav Republic of Macedonia, Serbia and Albania. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_100616.html

106. UNICEF. (2018). Evaluation of the Pilot MEND and RISE Cash Transfer Programmes in St. Kitts and Nevis. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Evaluation_of_the_Pilot_MEND_and_RISE_Cash_Transfer_Programmes_Barbados_2018-001.pdf

107. UNICEF. (2017). End line Evaluation of the Project for Improving Access to Quality Health and Education Services in the Northern and Upper East Regions of Ghana. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/REPORT_FINAL_082117.pdf

108. UNICEF. (2018). Evaluation of UNICEF’s Response to the Rohingya Refugee Crisis in Bangladesh. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_103442.html

109. UNICEF. (2018). Summative Evaluation of UNICEF Support for Education in Zimbabwe. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/UNICEF_Zimbabwe_Education_Evaluation_Report__2018-002.pdf

110. UNICEF. (2015). Final Evaluation Liberia Social Cash Transfer Programme. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Final_Evaluation_of_SCT_Pilot_Programme_Liberia.pdf

111. UNICEF. (2016). Impact Evaluation of Belize’s Conditional Cash Transfer Programme (BOOST). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/BOOST_Evaluation_report_final_Belize_2016-001.pdf

112. UNICEF. (2017). UNICEF Malawi’s Child-Friendly Schools Construction Component. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Child_Friendly_Schools_Evaluation_Report_.pdf

113. UNICEF. (2015). Final Evaluation Report: Youthlead Initiative: Most Significant Change and Lessons Learnt. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/YouthLEAD_Initiative_Final_Evaluation_Report_Most_Significant_Change_and_Lessons_Learnt_South_Sudan_2015-001.pdf

114. UNICEF. (2015a). Evaluation of UNICEF’s Peacebuilding, Education and Advocacy Programme (PBEA). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_PBEA-2015.html

115. UNICEF. (2015b). Evaluation of UNICEF-Supported MOET’s Initiative of Mother Tongue Based Bilingual

Education in Viet Nam 2006-2014. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_90456.html

116. UNICEF. (2017). Evaluation of the Process, Impact, Efficiency and Effectiveness of the Swaziland Child Friendly Schools (CFS) Programme - (INQABA). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Inqaba_evaluation_Final_report_edited_15_Jan_2018.pdf

117. UNICEF Evaluation Office. (2016). Communication for Development: An Evaluation of UNICEF's Capacity and Action: Synthesis Report. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/C4D_Evaluation_Report.pdf

118. UNICEF Evaluation Office. (2016a). Communication for Development: An Evaluation of UNICEF's Capacity and Action: Bangladesh Country Case Study. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/C4D_Evaluation_Report.pdf

119. UNICEF Evaluation Office. (2016b). Communication for Development: An Evaluation of UNICEF's Capacity and Action: Ethiopia Country Case Study. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/C4D_Evaluation_Ethiopia_Case_Study_2016-042.pdf

120. UNICEF Evaluation Office. (2018). Formative Evaluation of the Out-of-School-Children Initiative (OOSCI). UNICEF. Retrieved from https://www.unicef.org/evaldatabase/index_102939.html

121. UNICEF. (2017). Evaluacion Mejorando la Educación Básica de Niñas y Niños de la Amazonía y Sur Andino del Peru. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/2017.001_Eval_ACDI_Informe_Final.pdf

122. UNICEF. (2018). Outcome Evaluation of the Education Capacity Development Partnership Fund (CDPF) Phase I AND II: Final Report – Volume I - July 2017 – February 2018, Cambodia. Phnom Penh, Cambodia: UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/Cambodia_2018-002-Outcome-Evaluation-Education-CDPF-Volume-I.pdf

123. UNICEF. (2015). WCARO: Evaluation de la mise en oeuvre du projet francais de lutte contre les violences de genre en milieu scolaire. UNICEF. Retrieved from https://www.unicef.org/evaldatabase/files/WCARO_Rapport_Final_Eval_VGMS_2015-001.pdf

124. UNICEF. (2015). Evaluation of Integrated Post-Primary Education (IPPE) Pilot Project in Tanzania. UNICEF. Retrieved from [https://www.unicef.org/evaldatabase/files/IPPE_FINAL_REVISIED_VERSION_\(Tanzania_2015-009\).pdf](https://www.unicef.org/evaldatabase/files/IPPE_FINAL_REVISIED_VERSION_(Tanzania_2015-009).pdf)

125. UNICEF. (2017). My Village (ECD TV) Programme Evaluation Report UNICEF Lao PDR Country Office. Retrieved from <https://www.unicef.org/laos/media/296/file/UNICEF%20Lao%20PDR.pdf>

UNRWA

126. UNRWA. (2018). Evaluation of UNRWA's Engaging Youth project in Syria. UNRWA. Retrieved from https://evaluation.unrwa.org/sites/default/files/content/resources/resource_87481_35956_1540814515.pdf

UN WOMEN

127. Jimat Development Consultants. (2015). Final Evaluation of the United Nations Joint Programme for Gender Equality in Uganda. UN Women. Retrieved from <http://gate.unwomen.org/Evaluation/Details?evaluationId=4811>

WORLD FOOD PROGRAMME

128. Baker, J., Dunn, S., Vazquez, L., Kamah, Z., and Horst, N. (2016). Operation Evaluation. Liberia Protracted Relief and Recovery Operation 200550: Food Assistance For Refugees And Vulnerable Host Populations. World Food Programme. Retrieved from <https://documents.wfp.org/stellent/groups/public/documents/reports/>

wfp284812.pdf?_ga=2.96066716.1415253823.1551962475-647382046.1551962475

129. Bichard, A., Ngom, M., Thoreux, M., Watt, A., Leturque, H. (2018). Evaluation decentralisee de la modalite transfert monetaire utilisee dans le programme de cantines scolaires appuye par la PAM au Senegal. Rapport d'évaluation. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000074400/download/?_ga=2.95670428.1415253823.1551962475-647382046.1551962475

130. Cano, M., Murphy, C., Ahmed, F., and Suarez, H. (2018). Final Evaluation of McGovern-Dole-supported School Feeding Programme in Bangladesh. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000101229/download/?_ga=2.65190958.1415253823.1551962475-647382046.1551962475

131. Diagne, A., Solaroli, L. and Ba, A. (2017). PAA Africa programme in Senegal's Kédougou region: September 2013 - July 2016. World Food Programme. Retrieved from <https://docs.wfp.org/api/documents/WFP-0000022401/download/>

132. Downen, J., Sutter, P., Walters, T., Egeh, A., Kahsay, A., Alhaddad, R., and Vallet, M. (2016). Ethiopia Protracted Relief and Recovery Operation 200700 (2015-2018) Food Assistance for Eritrean, South Sudanese, Sudanese and Somali Refugees: An Evaluation of WFP's current operation and transition period. World Food Programme. Retrieved from https://documents.wfp.org/stellent/groups/public/documents/reports/wfp284322.pdf?_ga=2.62496556.1415253823.1551962475-647382046.1551962475

133. Dunn, S., and Otsola, J. (2017). Final evaluation of the World Food Programme USDA/McGovern-Dole International Food for Education and Child Nutrition Programme (FFE 615- 2013/041/00) in Kenya (Decentralized Evaluation). World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000063789/download/?_ga=2.205947220.652861801.1556650722-1217523547.1551967599

134. Dunn, S., Silvereano-Velis, J., Sin, S., Canteli, C., and Jahan, F. (2017). Endline Evaluation of USDA McGovern-Dole Grant Food for Education (FFE) Programme for WFP Cambodia 2013 - 2016. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000063575/download/?_ga=2.104585600.1415253823.1551962475-647382046.1551962475

135. Godden, K., Morris, M., Dunn, S., Merchant, D., and Horst, N. (2017). World Food Programme McGovern-Dole International Food for Education and Child Nutrition Programme (FFE 699-2013/036-00-B) in Liberia. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000023073/download/?_ga=2.138967347.1415253823.1551962475-647382046.1551962475

136. Goswami, B., Das, M., Singh, A., and Barooah, P. (2018). USDA Mc-Govern Dole FY14 End-Line Evaluation in Lao PDR. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000101261/download/?_ga=2.71415571.1415253823.1551962475-647382046.1551962475

137. Iswarawanti, D., Sulami, S., Ermayani, E., Habibie, Y., and Februhartanty, J. (2016). An Evaluation of the 2012-2015 Local Food-Based School Meal Programme. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000022110/download/?_ga=2.170973600.1415253823.1551962475-647382046.1551962475

138. Lister, S., Berhanu, D., Bjørnstad, L., Driscoll, Z., Fenton, G., Leach, A., and Sandford, J. (2019). Ethiopia: An evaluation of WFP's Portfolio (2012-2017). World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000102388/download/?_ga=2.242282756.701473542.1554592246-1217523547.1551967599

139. Lister, S., Girling, F., Bhatia, R., Mel, R., and Musoke, S. (2017). Sri Lanka: An evaluation of WFP's portfolio (2011-2015). World Food Programme. Retrieved from https://docs.wfp.org/api/documents/c01c87d9ec944cd39bd1cd7eef0084b1/download/?_ga=2.167476542.1415253823.1551962475-647382046.1551962475

140. Maunder, N., Hoogendoorn, A., Coombs, D., Fenton, G., and Carboni, L. (2017). South Sudan: An Evaluation of WFP's portfolio (2011-2016). World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000022545/download/?_ga=2.103994368.1415253823.1551962475-647382046.1551962475

141. Murphy, C., Fret, A., Baldeh, F., and Prieto, J. (2018). Gambia DEV 200327: Establishing the Foundation for a Nationally Owned Sustainable School Feeding in the Gambia. From 2012 to 2017. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000073914/download/?_ga=2.103003520.1415253823.1551962475-647382046.1551962475
142. Peacocke, B., Tadesse, S., and Moshoeshoe, R. (2018). Evaluation of the National School Feeding Programme in Lesotho, in consultation with the Lesotho Ministry of Education and Training 2007-2017. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000073908/download/?_ga=2.163156156.1415253823.1551962475-647382046.1551962475
143. Torres, S., Leguene, P., Riveros, L., Cuartas, M., Tellería, R., Rodriguez-Leon, S., and Murgueitio, C. (2018). Evaluación Final del Programa de País del PMA en el Estado Plurinacional de Bolivia - PP200381 2013-2017. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000103099/download/?_ga=2.104439168.1415253823.1551962475-647382046.1551962475
144. Turner, S., Bossuyt, A., Leather, C., Mhina, E., Driscoll, Z., and Mahali, A. (2015). Country Portfolio Evaluation. United Republic of Tanzania: An evaluation of WFP's portfolio (2011–2014). World Food Programme. Retrieved from https://documents.wfp.org/stellent/groups/public/documents/reports/wfp278942.pdf?_ga=2.260796557.1415253823.1551962475-647382046.1551962475
145. Visser, M., Pinto, J., Garrido, C., Leturque, H. (2016). Evaluation D'Operation Sao Tomé at Principe - Project de Developpement DEV 200295- Transitions ver un Programme National d'Alimentation et de Santé Scolaire a Sao Tomé et Principe 2012-2016: Evaluation d'operation de PAM (2012-2016). World Food Programme. Retrieved from https://documents.wfp.org/stellent/groups/public/documents/reports/wfp282990.pdf?_ga=2.167476542.1415253823.1551962475-647382046.1551962475
146. Visser, M., Alder, D., Bhatia, R., Bultosa, G., Berhanu, D., and Fenning, C. (2018). Final Evaluation of WFP'S USDA McGovern-Dole International Food for Education and Child Nutrition Programme in Afar and Somali Regions in Ethiopia from 2013 to 2017. World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000101695/download/?_ga=2.100783873.1415253823.1551962475-647382046.1551962475
147. Webb, K., Kamlongera, A., Makoka, D., Sopo, B., and Dunn, S. (2018). Final Evaluation of the School Meals Programme in Malawi with support from the United States Department of Agriculture, and the Governments of Brazil and the United Kingdom (2013-2015). World Food Programme. Retrieved from https://docs.wfp.org/api/documents/WFP-0000074165/download/?_ga=2.123976842.1415253823.1551962475-647382046.1551962475

