

Academic Learning Measurement and Assessment Tools in Education in Emergencies: Identifying, Analyzing, and Mapping Tools to Global Guidance Documents

Kate Anderson, Lindsay Read, and Elena Losada
Unbounded Associates

Final report prepared for the Inter-agency Network for Education in Emergencies
Education Policy Working Group, Quality and Equitable Learning Outcomes Workstream

Submitted March 3, 2020



Acknowledgements

The authors acknowledge Margi Bhatt of the Inter-agency Network for Education in Emergencies and the members of the QELO work stream, especially Gerhard Pulfer of Porticus, Audrey Nirrengaarten of Agence Française de Développement (AFD), and Anjuli Shivshanker of USAID, for their support in developing this report. We also greatly appreciate the thought partnership of the EASEL Lab at Harvard University, through the leadership of Sonya Temko, Rebecca Bailey, and Stephanie Jones. We also would like to thank Luis Crouch and Seamus Hegarty for peer reviewing a draft of this paper. Finally, we are grateful to the following individuals for sharing information about their programs and tools: Roxane Caires of 3EA, Mary Sugure of EDC, Julius Kinyash of the Kakuma refugee camp, Antoine Marivin of SEA-PLM, Silvia Montoya and Adolfo Gustavo Imhof of the UNESCO Institute for Statistics, Clara Pava of Save the Children, and Maria Layhing Siu Munro of ArmandoAid.

Table of Contents

Executive Summary	7
Part 1. Background	23
The global landscape	23
INEE’s work to date	24
Assessing learning outcomes in EiE contexts	24
Methodology	26
Parameters and taxonomy	26
Selection criteria	28
Part 2. Overview of Tools, Frameworks, and Program Approaches	32
Guidance documents	32
Assessment tools	34
Validity and reliability	34
Inclusion of children in emergency contexts	39
Target populations	40
Administration of assessment	43
Accessibility and costs	44
Assessment construction and purpose	44
Contextual information collected	44
Program approaches	47
Part 3. Domain Mapping: Literacy and Numeracy	51
Coding system	51
Mapping literacy domains and sub-domains	52
Mapping numeracy domains and sub-domains	56
National examinations and classroom assessments	60
Part 4. Analysis and Conclusions	63
Application of guidance documents and measurement tools for EiE contexts	63
Connections between local, national, regional, and international tools and frameworks	64
Summary	66
Bibliography	68
Annex A. List of Stakeholders Consulted	74
Annex B. Guidance Documents	75
Annex C. Measurement Tools	78
Annex D. Program Approaches	86
Annex E. Descriptions of Domains and Sub-Domains	93

List of figures

Figure 1. Guidance documents, program design approaches, and measurement and assessment tools ...	8
Figure 2. Age and grade ranges of tools included in study	11
Figure 3. Relationship between guidance documents, program design, measurement/assessment tools, and monitoring and evaluation.	27
Figure 4. Countries of interest for analysis of measurement tools	31
Figure 5. Age and grade ranges of tools included in study	41
Figure 6. Literacy domains and sub-domains.....	53
Figure 7. Numeracy domains and sub-domains.....	53
Figure 8. Literacy sub-domains by number of tools.....	55
Figure 9. Numeracy sub-domains by number of tools.....	58
Figure 10. Content descriptors for reading.....	93
Figure 11. Content descriptors for mathematics	94

List of tables

Table 1. Targeted grade or age range of tools used in countries where at least one tool has been administered in the last five years	12
Table 2. Summary of measurement tool characteristics	14
Table 3. Types of assessment tools by literacy domain and sub-domain.....	20
Table 4. Types of assessment tools by numeracy domain.....	21
Table 5. Education levels covered in the mapping exercise	30
Table 6. Reading minimum proficiency levels—UNESCO Institute for Statistics.....	32
Table 7. Mathematics minimum proficiency levels—UNESCO Institute for Statistics	33
Table 8. Types of psychometric properties	35
Table 9. Characteristics of the populations, reliability and validity of assessment tools	36
Table 10. Targeted grade or age range of tools in countries where at least one tool has been administered in the last five years	42
Table 11. Summary of measurement tool characteristics	45
Table 12. Types of assessment tools by literacy domain and sub-domain.....	54
Table 13. Types of assessment tools by numeracy domain.....	57
Table 14 Conflict- or crisis-affected GPE grant recipients prioritizing learning assessment	65
Table 15. Global frameworks included in analysis	75
Table 16. System-monitoring measurement tools included in mapping exercise with internationally/regionally- comparable results	78
Table 17. Multi-purpose measurement tools included in mapping exercise where results cannot be compared internationally.....	81
Table 18. Program approaches included in the mapping exercise	86

List of Boxes

Box 1. Example EiE case: Administration of ASER, EGMA, and EGRA in Syria	50
Box 2. Classroom assessments for Afghan refugees in Central Greece.....	62

Acronyms

3EA— Education in Emergencies: Evidence for Action

AFD— Agence Française de Développement

ASER—Annual Status of Education Report

CLAs—Citizen-led assessments

ECW— Education Cannot Wait

EDC—Education Development Center

EGMA —USAID Early Grade Mathematics Assessment

EGRA—USAID Early Grade Reading Assessment

EGWA—Early Grade Writing Assessment

EPWG—Education Policy Working Group

GAML—Global Alliance to Monitor Learning

GPE—Global Partnership for Education

HALDO—Save the Children’s Holistic Assessment of Learning and Development

ICCS— International Civic and Citizenship Education Study

IDELA—Save the Children’s International Development and Early Learning Assessment

INEE—International Network of Education in Emergencies

IRC—International Rescue Committee

ISCED—International Standard Classification of Education

LaNA— International Association for the Evaluation of Educational Achievement’s Literacy and Numeracy Assessment

LLECE— Latin American Laboratory for Assessment of the Quality of Education

LSCE MENA—Life Skills and Citizenship Education in Middle East and North Africa

MELQO—Measuring Early Learning Quality and Outcomes

MELQO-MODEL— Measuring Early Learning Quality and Outcomes, Measure of Development and Early Learning Module

MENA—Middle East and North Africa

MICS-ECDI— Multiple Indicator Cluster Survey, Early Childhood Development Index

MICS-FLS—Multiple Indicator Cluster Survey, Foundational Learning Skills

MLA—UNRWA Monitoring Learning Achievement

NAFs—national assessment frameworks

NYU—New York University

OCHA—Office for the Coordination of Humanitarian Affairs (OCHA)

OLA—Out-of-School-Youth Literacy Assessment and eOLA

PASEC— Programme d’Analyse des Systèmes Éducatifs de la CONFEMEN

PILNA— Pacific Island Literacy and Numeracy Assessment

PIQ—program implementation quality

PIRLS—Progress in International Reading Literacy Study

PISA-D—Program for International Student Assessment for Development

QELO—Quality and Equitable Learning Outcomes

REAL— Research for Equitable Access and Learning, University of Cambridge

SDGs—Sustainable Development Goals

SEACMEQ—Southern and Eastern Africa Consortium for Monitoring Educational Quality

SEA-PLM—Southeast Asia Primary Learning Metrics

SEL—social and emotional learning

STAR—World Vision’s School-based Test About Reading

TERCE—Third Study of Assessment of the Quality of Education

TIMSS—Trends in International Mathematics and Science Study

UIS—UNESCO Institute for Statistics

UNESCO IBE—UNESCO International Bureau of Education

UNESCO IIEP— UNESCO International Institute for Educational Planning

UNHCR—UN High Commissioner for Refugees

USAID—United States Agency for International Development

Executive Summary

Project Background

A sector-wide shift away from a focus on school access and attendance toward equity and learning will require common understanding among implementers and funders working in the international education in emergencies (EiE) sector and between the EiE and development sectors, and alignment of their measurement tools. Recognizing this, the Quality and Equitable Learning Outcomes (QELO) work stream within the Inter-agency Network for Education in Emergencies (INEE) Education Policy Working Group (EPWG) commissioned two separate mapping exercises of academic and social and emotional learning (SEL) program approaches,¹ measurement tools, and monitoring and results frameworks, the aim being to map what quality education in emergencies constitutes at the program and individual learning levels, to inform the linkages with global measurements.

This report contributes to the academic component of the QELO work stream by outlining the landscape of tools, guidance documents, and program approaches currently used to measure learning outcomes in EiE contexts, and by providing an analysis of the links and alignment among these in order to inform a more streamlined measurement of learning outcomes in emergencies.

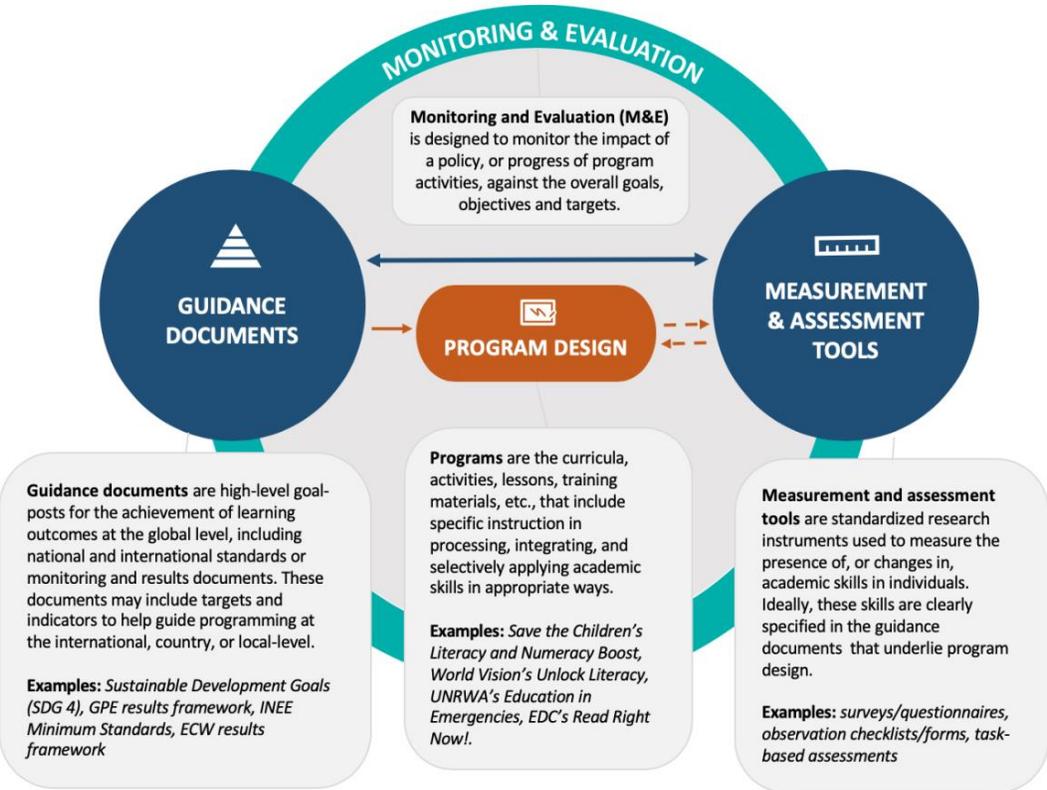
Part 1 of the report provides an overview of tools, frameworks, and program approaches and compares each assessment's purpose, design, target population, and various other parameters. Part 2 analyzes commonalities and differences across the domains and sub-domains assessed by each tool. Part 3 summarizes the links across the measurement tools, frameworks, and program approaches, and their application in EiE contexts.

Overview of Methodology

This report reviews 31 measurement tools (23 distinct approaches, with 8 citizen-led assessments [CLAs] comprising the remaining tools), 8 guidance documents, and the program approaches of 4 organizations that were used in at least 1 country affected by conflict or crisis in the past 5 years. Selection of these tools was based on three key factors: coverage, quality, and relevance. We located the tools and frameworks through desk research, conversations with key stakeholders, and an online survey of INEE members. To enable comparisons across the various tools, we adopted the framework and coding scheme created by the UNESCO International Bureau of Education (UNESCO IBE) and the UNESCO Institute for Statistics (UIS) to map content domains and sub-categories in mathematics and reading, and to identify commonalities and disparities across content areas and coverage in the measurement of academic skills.

¹ The SEL mapping was conducted by the EASEL Lab at the Harvard Graduate School of Education.

Figure 1. Guidance documents, program design approaches, and measurement and assessment tools



The tools included cover children and youth ages 3 to 18 who are living in conflict-affected contexts around the globe, including Africa (Francophone, Anglophone, Lusophone), South Asia, Latin America, the Pacific Islands, Eastern Europe, Central Asia, and the Middle East.

Figure 2 illustrates the development stages targeted by each of the tools.

Key Findings

The key findings described below are structured to respond to the priority questions of the INEE Quality and Equitable Learning Outcomes work stream.

What considerations or barriers prevent including crisis- and conflict-affected populations in assessments?

The humanitarian sector faces a number of unique challenges in delivering and measuring learning in emergency contexts. Moreover, without a consistent set of goals as to what children should know and be able to do, it is difficult to build a functional learning assessment system:

- Existing assessment systems either are not set up to capture populations on the move in crisis contexts, or they respond inadequately to additional strain such emergency situations put on weak systems. This results in a fragmented system with non-comparable data and few or no baseline reference points. Moreover, when displaced children are integrated into national systems, they are not registered as “refugees” or “IDPs.” While this may be a positive step in terms of reducing stigma, it makes it impossible to track these children’s progress over time and makes it difficult for policymakers to address their specific needs.
- Relatedly, this fragmentation can be amplified where parallel education systems exist. For instance, some refugees attend the public schools in their host community while others reside in camps where education services are provided by the international community or local NGOs.
- National governments, particularly those of countries affected by conflict and crisis, are often unable to collect data and manage assessment systems. In crisis contexts, the short-term humanitarian aid provided for education can limit the implementing agencies’ ability to help strengthen national and local institutions’ capacity to collect data.
- In some situations, particularly where refugee or displaced populations are not integrated into the local education system, such as the Rohingya refugee crisis in Bangladesh, neither the host country nor the country of origin allows their national curriculum to be used to educate refugee children (McPherson & Paul, 2019).
- Children and youth who have faced conflict and crisis may bear a cognitive load that is far heavier than that of their peers. This means that conflict- or crisis-affected children would be at a disadvantage in taking standardized tests that have been validated with populations not facing similar difficulties. Traditional test protocols also may not be appropriate.

Several assessment tools were developed specifically for EiE contexts, namely Save the Children’s Holistic Assessment of Learning and Development (HALDO), the UNRWA Monitoring Learning Achievement (MLA), and the Education Development Center (EDC) Out-of-School-Youth Literacy Assessment (OLA). Save the Children’s International Development and Early Learning Assessment (IDELA) also included children in EiE settings in its pilot. However, other assessment tools used in a range of contexts do not

disaggregate or adequately sample children affected by conflict or crisis. Of the 15 countries affected by conflict or crisis whose reading data appear on the USAID Early Grade Reading Barometer (EGRA) website, only 4 mention the emergency context in the EGRA reports.² None of the reports disaggregated data by refugee or IDP status or noted efforts to include children affected by conflict and crisis in the study. None of the reports stated that the samples included EiE contexts or displaced student populations.

The exclusion of EiE contexts is evident in regional and international assessments as well. The 2005 Southern and Eastern Africa Consortium for Monitoring Educational Quality (SEACMEQ) report for Uganda notes that “schools in areas affected by serious military conflicts and special schools were excluded” (Byamugisha & Ssenabulya, 2005, pp. 33-34) from the study. Data collection for Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC) was also discontinued in conflict-affected areas in Mali in 2012-2013, and these areas were excluded from their 2014 study.

How, where, and with whom are academic measurement tools used?

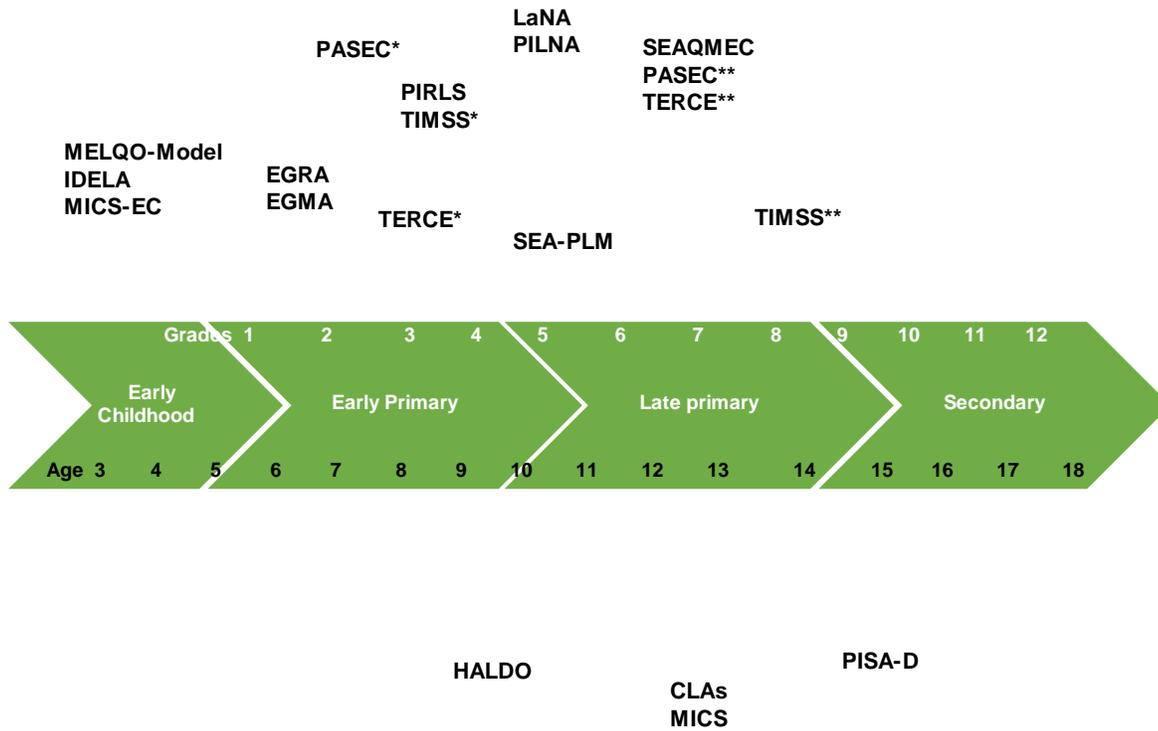
We conducted a mapping of cross-national (used in more than one country) academic learning assessment tools in 61 countries where EiE is provided. Of the 61 countries, only 2—Eritrea and Libya—did not participate in any of the cross-national tools. Some tools are international (covering multiple countries and more than one region): the MICS surveys, TIMSS, PIRLS,³ EGRA, EGMA, MELQO, and IDELA. Others cover specific regions: PASEC, SEACMEQ, TERCE, PILNA, and SEA-PLM target Francophone Africa, Southern and Eastern Africa, Latin America, the Pacific Islands, and Southeast Asia, respectively. Several more recently developed tools, PISA-D, HALDO, and LaNA, have been piloted in a few countries with the intention to expand coverage.

The assessment tools we selected for coding cover a range of target populations from early childhood to age 18. Annex C contains descriptions of the 30 multi-country assessments included in this exercise. MELQO, IDELA, and MICS-ECDI target early childhood; EGRA and EGMA target the early primary grades; and several tools, including PASEC, TERCE, and TIMSS, have an early grade assessment and another they use at the end of primary school. SEACMEQ, LaNA, PILNA, and SEA-PLM target the last grades of primary school. Several tools that assess children outside of school (including MICS, HALDO, and the CLAs) administer the same assessment to children within a wide age range (e.g., from 6 to 18 years).

² Countries with data on earlygradereadingbarometer.org that are considered affected by conflict or crisis for this report include Bangladesh, Cambodia, DRC, Ghana, Iraq, Jordan, Kenya, Liberia, Malawi, Mali, Nigeria, Palestine, Papua New Guinea, The Philippines, and Uganda.

³ PIRLS is primarily used by high-income countries, and only two middle-income countries affected by conflict in our list of countries of interest, Colombia (2011) and Turkey (2001), have participated in the last few rounds.

Figure 2. Age and grade ranges of tools included in study



How an assessment is administered varies according to context, purpose, and target population. Tools designed for children under grade 3, CLAs, assessments embedded in MICS household surveys, and tools designed specifically for displaced or out-of-school children are usually orally delivered face-to-face by a trained assessor from the administering agency. These occur in a home, school, or community center, depending on the tool and its purpose. CLAs are usually administered by trained citizen volunteers in the home environment, while the PISA-D, PIRLS, PASEC, and others that are administered to older children in school are typically paper based and include multiple-choice and open-response questions.

In terms of assessment construction, the large majority (80%) of tools included in the mapping are competency based, meaning that they test specific skills. Six of the tools are content based, so they are constructed with a specific curriculum content in mind. Several regional assessments a few others are content based, including SEACMEQ, TERCE, and SEA-PLM, and TIMSS.

Twelve of the thirty-one tools are designed for system monitoring, which means they yield results that are comparable both year-after-year and cross-nationally. These include two learning modules of the MICS household surveys and international and regional assessments. The remaining 11 tools yield results that are context specific, multi-purpose, and not comparable internationally. These include some of the early childhood and early grade assessments such as EGRA and EGMA, HALDO, and the CLAs. Table 1

indicates the system-monitoring and multi-purpose tools present at each grade level in the focus countries.

Specific program approaches often tweak existing tools to create bespoke assessments that cater to the unique needs of each program. The Literacy Boost assessment tool is similar in many ways to the EGRA tool and Numeracy Boost to EGMA. World Vision’s School-based Test About Reading (STAR) tool also draws from EGRA and the Literacy Boost toolkit, but it asks students to identify fewer words and excludes the listening-comprehension component. EDC adapted the paper-based EGRA to an online format (the eOLA) specifically to capture information on older out-of-school youth; it was adapted from the OLA, which was better suited to the need for rapid information.

Table 1. Targeted grade or age range of tools used in countries where at least one tool has been administered in the last five years

	System-Monitoring Tools				Multi-Purpose Tools		
	Early childhood	Grades 2 or 3	Late primary	Secondary	Early childhood	Grades 2 or 3	Age range*
Afghanistan					⊙	⊙	
Angola			⊙			⊙	
Bangladesh	⊙				⊙	⊙	⊙
Benin	⊙	⊙				⊙	⊙
Burkina Faso		⊙					
Burundi						⊙	
Cambodia			⊙	⊙		⊙	
Cameroon	⊙	⊙			⊙		
Chad	⊙	⊙				⊙	⊙
CAR	⊙	⊙				⊙	⊙
Colombia		⊙	⊙	⊙	⊙		
Congo, DRC	⊙	⊙				⊙	⊙
Côte d'Ivoire	⊙	⊙					
El Salvador	⊙	⊙	⊙	⊙			⊙
Eritrea							
Ethiopia					⊙	⊙	
Gambia	⊙						⊙
Ghana	⊙		⊙		⊙	⊙	⊙
Guatemala		⊙	⊙	⊙	⊙	⊙	
Guinea	⊙	⊙					
Guinea-Bissau	⊙						⊙
Haiti					⊙	⊙	
Honduras	⊙	⊙	⊙	⊙	⊙	⊙	⊙
Indonesia	⊙		⊙	⊙	⊙	⊙	
Iraq	⊙					⊙	⊙
Jordan			⊙	⊙	⊙	⊙	
Kenya	⊙		⊙		⊙	⊙	⊙
Kyrgyzstan	⊙				⊙	⊙	⊙

	System-Monitoring Tools				Multi-Purpose Tools		
	Early childhood	Grades 2 or 3	Late primary	Secondary	Early childhood	Grades 2 or 3	Age range*
Lebanon	○		○		○		○
Lesotho	○		○		○		○
Liberia					○	○	
Libya							
Malawi	○		○		○	○	○
Mali	○	○			○	○	○
Mauritania	○						
Mozambique			○		○	○	○
Myanmar			○		○		
Nicaragua		○	○		○	○	
Niger		○			○	○	
Nigeria	○					○	○
Pakistan	○			○	○	○	○
Palestine	○	○	○		○	○	○
Papua New Guinea			○		○	○	
Philippines			○	○	○	○	
Senegal	○	○		○	○	○	○
Sierra Leone	○				○	○	○
Solomon Islands			○		○		
Somalia	○				○	○	
South Sudan	○					○	
Sudan	○				○		○
Syria			○			○	
Togo	○	○					○
Tonga	○		○			○	○
Tunisia	○		○				○
Turkey			○	○			
Uganda			○		○	○	○
Ukraine	○		○				
Venezuela		○	○				
Yemen						○	
Zambia			○	○	○	○	
Zimbabwe	○		○		○	○	○

*This generally includes children of primary and secondary school age, but it is not defined by grade because tools are administered outside of the classroom. CLAs, the MICS Foundational Skills Module, and HALDO are administered this way; they use the same tool for ages 4 to 16.

Which features of children’s context, background, and experiences are being considered?

Most tools gather contextual and background information on students, families, schools, administrators and/or teachers, along with assessment results. All tools at the least collect information on student characteristics such as gender, socioeconomic status, language, and education level. Having information about particular characteristics in different settings provides a more complete understanding of potential learning challenges, or successes, that can guide programming and other decisions. Some tools collect information on student learning outside of school (e.g., time spent reading or studying at home, extracurricular activities), household characteristics (e.g., parental education, mother tongue, socioeconomic status, use of information and communications technology at home), school characteristics (e.g., school infrastructure and resources, management and leadership, teacher absenteeism, homework), and village/community facilities (e.g., remoteness, health centers, electricity).

Tools such as PILNA, EGRA, PIRLS, TIMSS, and TERCE often include background questionnaires for parents, teachers, and principals that are administered along with the learning assessments for children. These questionnaires help provide a more robust picture of the education landscape in each context. For example, the learning assessment modules that are part of MICS household surveys collect a host of background information on household and family characteristics, including location, health, wealth, and disabilities, which can promote robust learning analysis. However, these tools have not been used to collect information about refugee or IDP status. Efforts have been made only recently to achieve consensus on how to collect statistics on refugees and IDPs, including by UIS and its partners (UIS, FHI360, Oxford Policy Management, & REAL, 2018), the UN High Commissioner for Refugees (UNHCR, 2018), and the Education Equity Research Initiative (Soares, Smiley, & Lavan, 2018).

In terms of program approaches, the measurement tools employed by Save the Children—the Literacy Boost Assessment, Numeracy Boost Assessment, and IDELA—measure the skills of all students in a particular program using a continuous scoring system that allows for equity analysis. Paired with surveys that gather information on student background and school context, this enables programs to identify and respond to the unique needs of disadvantaged groups.

Table 2. Summary of measurement tool characteristics

Assessment	Purpose	Construction	Availability	Method	Site	Context
PASEC	System monitoring	Competency based	Country fee for participation and implementation	Grade 2 oral; grade 6 written multiple choice	School	Background questionnaire includes student, household, and school characteristics
PILNA	System monitoring	Competency based	Country fee for participation and implementation	Written multiple choice	School	Background questionnaire for student, teacher, and principal, including information on school resources and management
PIRLS	System monitoring	Competency based	Country fee for participation and implementation	Written multiple choice and open response	School	Background questionnaires cover student, household, and school characteristics

Assessment	Purpose	Construction	Availability	Method	Site	Context
PISA-D	System monitoring	Competency based	Country fee for participation and implementation	School tool is written multiple choice and open response; out-of-school tool is tablet based	School; Home (Out of school component)	Background questionnaires cover student, home, school, and community characteristics
MICS-ECDI	System monitoring	Competency based	Open source	Oral	Home	Comprehensive background information collected as part of general MICS survey
MICS-Foundational Learning Skills	System monitoring	Competency based	Open source	Oral	Home	Comprehensive background information collected as part of general MICS survey. Information collected in the module includes reading habits and home language.
LaNA	System monitoring	Content based	Not open source	Written multiple choice	School	Background information on school and household characteristics
UNRWA MLA	System monitoring	Content based	Not open source	Written multiple choice and open response	School	Background questionnaires on classroom and school characteristics
SEA-PLM	System monitoring	Content based	Country fee for participation and implementation	Written multiple choice and open response	School	Proposal for background information includes student, household, school, and community characteristics
SEACMEQ	System monitoring	Content based	Country fee for participation and implementation	Written multiple choice and open response	School	Background information on student, household, and school characteristics; access to books and tutoring
TERCE	System monitoring	Content based	Country fee for participation and implementation	Written multiple choice and open response	School	Background questionnaires for teachers, principals, students, and families, including characteristics like school management
TIMSS	System monitoring	Content based	Country fee for participation and implementation	Written multiple choice and open response; Online tablet-based option in 2019	School	Background information on school, household, and student characteristics; attitudes toward learning
OLA	Multi-purpose	Competency based	Not open source	OLA is oral; eOLA uses digital tablets	Nonformal education programs	Demographic information, information on reading, nonformal and formal education background, literacy goals
Literacy Boost	Multi-purpose	Competency based	Not open source	Oral	School	Background information on household and school characteristics; reading habits at home
Numeracy Boost	Multi-purpose	Competency based	Not open source	Oral	School	Information about students' exposure to math outside of school

Assessment	Purpose	Construction	Availability	Method	Site	Context
MELQO	Multi-purpose	Competency based	Open source	Oral	School, home, or informal learning center	No standardized collection of background information
STAR	Multi-purpose	Competency based	Open source	Structured task	In or near school	Recommended background includes demographic and economic information, home literacy environment
CLAs*	Multi-purpose	Competency based	Open source; fee for implementation	Oral	Home	Comprehensive background information on demographics, and school, community, child, and household characteristics
EGMA	Multi-purpose	Competency based	Open source; fee or own cost for implementation	Oral	School	No standardized collection of background information
HALDO	Multi-purpose	Competency based	Open source	Oral	School, home, or informal learning center	Background information on demographics, home learning environment, language
IDELA	Multi-purpose	Competency based	Open source but must sign MOU	Oral	Varies	Information collected on parent education, enrolment in ECE, household assets
EGRA	Multi-purpose	Competency based	Open source; fee or own cost for implementation	Oral	School	No standardized collection of background information

In what contexts have measurement tools been validated?

According to the tool developers, all measurement tools assessed in this study have been designed and piloted to determine some psychometric standards of validity and reliability. However, the transparency of the results and the extent to which validity and reliability have been established varies. The tools are typically validated on an as-needed basis as part of the adaptation process when used in a new context or language.

For tools that are to be used in various contexts and across multiple languages, such as EGRA and EGMA, detailed toolkits provide instructions on how to adapt the instrument. Three tools—HALDO, the MLA, and OLA—have been validated in an EiE setting. For instance, HALDO was piloted on 852 children from 27 centers and schools in the Dadaab refugee complex in Kenya (Krupar, D’sa, Westrope, & Johna, 2019). The study found strong evidence of reliability and validity for identifying the baseline learning of 4- to 12-year-olds; however, the tool still requires contextualization to local social and cultural norms when adapted to other locations. Part 1 of the report includes additional information on the populations each tool was developed for and piloted with, and a brief summary of reliability and validity.

How is assessment applied in EiE contexts?

Several examples of assessments in EiE contexts were provided by INEE members. In Syria, for instance, the War Stressor Survey, which measures exposure to conflict and the emotional repercussions of trauma, and the Snapshot of School Management and Effectiveness were administered along with the EGRA and EGMA learning assessments in order to provide a full picture of primary education in opposition-held areas in Syria. The simultaneous information on academic performance, student well-being, and the insights of school staff members supports an evidence-based design of donor programming (see **Box 1**).

Another example is the Oinofyta Community School, a project of the NGO ArmandoAid that was founded to serve Afghan children in a refugee camp in Central Greece. The school curriculum included literacy, numeracy, and several other subjects. Classroom assessments were used to monitor learning, and each child's behavior and general performance. The information on learning, which was used to improve teaching and operation of the school, was shared with donors and other development partners to highlight both progress and challenges (see **Box 2**).

What are the drivers of cost?

MICS-based tools, CLAs, MELQO, EGRA, EGMA, and IDELA are open-source tools that are available online, but there are costs associated with in-country implementation of all of them. For CLAs, the implementation cost is around US\$200,000, depending on how the assessment is implemented (i.e., national or sub-national). All regional assessments and a few international assessments require countries to pay a participation fee in addition to the cost of implementing the assessment in country. The cost of regional assessments (e.g., TERCE, SEACMEQ, PASEC), including access to the tool and in-country implementation, ranges from US\$200,000 to US\$500,000 per country, depending on the assessment program and local costs. For PISA-D, PIRLS, and TIMSS, the cost is about US\$800,000, which may vary according to the case (UIS, 2018a). The TIMSS and PIRLS program fees are US\$225,000. PISA program fees are around the same, but the exact fees vary by country and are related to how much a country pays in OECD membership fees. Fees for regional assessment also vary in accordance with a country's economic status and ability to pay.

Test administration represents about half of the total cost of these major international assessments, which is largely driven by the cost of field testing and supervision; institutional costs, driven by personnel costs and fees, account for around 25%. Other costs include test preparation, processing, analysis, and dissemination of findings (UIS, 2016b). The cost of administering EGRA is much lower overall and is determined primarily by the cost of field testing and personnel (Wagner, 2011).

How are measurement tools aligned to guidance documents?

There are some linkages among international, regional, national, and local assessment frameworks and tools. As the agreed-to global reference point, the Sustainable Development Goals (SDGs) and related

indicators are widely influential, and many regional and organization-specific frameworks on learning align with SDG guidance. Some regional frameworks define themselves as a regional extension of the SDGs, thereby owning them and adapting them to their regional aspirations. Several regional frameworks, like the Continental Strategy for Education in Africa framework, use indicators that are identical to the Sustainable Development Goal indicators to measure learning.

According to the UIS, nine cross-national learning assessments meet the criteria to measure SDG4, indicator 4.1.1, the proportion of young people achieving minimum proficiency in reading and mathematics in grades 2 and 3, at the end of primary school, and at the end of secondary school. These agreed-to assessments are LaNA, PASEC, PILNA, PIRLS, PISA, SEACMEQ, SEA-PLM, TERCE, and TIMSS. Although these are listed as the official assessments for measurement, a new methodology to monitor SDG4.1.1 progress that involves agreement on minimum proficiency levels allows CLAs and national assessments to be included in the SDG4 monitoring (USAID & UIS, 2019).

In the EiE sector, regional and NGO documents reference common standards set out by the INEE Minimum Standards for the assessment of learning outcomes; however, the INEE Minimum Standards do not recommend, require, or align with a specific learning assessment. The LSCE MENA framework points to specific skills outlined in the INEE Minimum standards. Education Cannot Wait (ECW) uses the INEE Minimum Standards as part of the benchmark for instruction and learning processes across ECW-supported programs.

There is also evidence that some program approaches deliberately link to regional and national assessments and standards. NGOs implementing learning programs use national, regional, and international assessments for monitoring and measurement. For example, when FHI360 implements accelerated education, SEL, and workforce-preparation programming, they include learning outcome assessments like the Annual Status of Education Report (ASER), EGRA, and EGMA, depending on what is needed and relevant for the planned intervention. Other NGOs with the necessary technical capacity develop new curriculum and assessment tools for education interventions. Save the Children, for instance, has developed and administered the HALDO, IDELA, Literacy Boost, and Numeracy Boost assessments as skills diagnostics, for monitoring and comparing the effectiveness of national interventions, impact evaluation, and program evaluations.

Which academic competencies are prioritized among the different tools?

While all assessments included in our analysis measure some understanding of basic math and/or literacy, each tool varies in terms of the particular skills, knowledge, and competencies children are required to demonstrate. To enable comparisons across the different tools, we adopted the framework and coding scheme created by the UNESCO IBE and the UIS to map content domains and sub-categories in literacy and math. There are three literacy domains—reading, linguistic, and metalinguistic competencies—and six math domains—math proficiency, number knowledge, measurement, statistics, geometry, and algebra. The additional domain of writing was included, although it is not part of the UNESCO IBE and UIS coding scheme.

Of the literacy domains, reading competency is the most prevalent, and all tools reviewed include some form of decoding and reading comprehension sub-domains. Reading and comprehending a simple text are understood across the globe as the fundamental skills children need to succeed in other areas of education. As children pass through the grade levels, more and more academic content is transmitted to them through text, and their ability to learn depends on their ability to read. Reading is also an essential skill for self-guided learning outside the classroom. As would be expected, beginner tasks such as decoding are most frequently found in assessments in early childhood and the early grades, while more complex tasks such as reading comprehension are found most frequently in assessments in the higher grades.

Vocabulary is the most common sub-domain within the linguistic competency domain, with about 15 tools testing some form of vocabulary knowledge. It is most commonly tested with tools that target younger populations, like basic word identification using images in the early childhood tools, and with the regional assessments and EGRA that test a more comprehensive set of reading sub-domains. Speaking and listening are directly assessed with five and six tools, respectively, like EGRA and some program approaches, but they are rarely included in the regional and international written tests. All the tools that capture speaking and listening are administered orally.

Twelve tools assess metalinguistic competency. The subset of tools that assess phonological awareness are similar to those that assess vocabulary, like tools targeting the early years, regional assessments, and comprehensive literacy tools such as EGRA.

Writing is targeted in some of the regional assessments, as well as PIRLS, EGRA, and two early childhood tools, IDELA and MELQO. An additional writing component is being developed for several tools.

Table 3. Types of assessment tools by literacy domain and sub-domain

Category	Tools	Reading		Linguistic			Metalinguistic	Writing
		Decoding	Comprehension	Listening	Speaking	Vocabulary	Phonological awareness	
System-monitoring tools								
International assessments	PISA-D, TIMSS, PIRLS, LaNA		Darkest blue			Lightest blue	Lightest blue	Lightest blue
Regional assessments	LLECE, SEACMEQ, PASEC, PILNA, SEA-PLM, UNRWA MLA	Lightest blue	Darkest blue	Lightest blue	Lightest blue	Darkest blue	Darkest blue	Lightest blue
Household	MICS-ECDI, MICS-FLS	Darkest blue	Lightest blue			Lightest blue	Lightest blue	
Multi-purpose tools								
Foundational skills ⁴	MELQO, EGRA, STAR, Literacy Boost	Darkest blue	Lightest blue	Lightest blue	Lightest blue	Lightest blue	Lightest blue	Lightest blue
Citizen led	ASER, UWEZO, LearnNigeria, etc.	Darkest blue	Darkest blue					
Designed for EIE contexts	HALDO, IDELA, OLA	Darkest blue	Lightest blue	Lightest blue	Lightest blue	Darkest blue	Lightest blue	Lightest blue

-  Darkest blue indicates that all tools in the category include sub-domain
-  Lightest blue indicates that half or more tools in the category include sub-domain
-  Lightest blue indicates that one or a few tools in the category include sub-domain
-  No color indicates that no tools in the category include sub-domain

No tools cover all of the sub-domains in mathematics, and tools that cover more advanced content, such as vectors and probability, are particularly rare. However, all tools, regardless of their intended use, measure number knowledge. PISA-D covers the greatest breadth of mathematics material, measuring all sub-domains except for the most basic concepts (e.g., pre-number ideas) and one of the most advanced concepts (i.e., vectors). Other regional and international tools that have broad coverage in mathematics are TIMSS, TERCE, SEACMEQ, and SEA-PLM. Apart from number knowledge, content in these tools generally includes measurement, basic data tables, problem-solving, and geometry.

Unlike the longer school-based assessments, rapid assessments like HALDO and CLAs focus solely on numeracy. For these kinds of tools, numbers and number systems are the only sub-domains included. In HALDO, the numbers task involves asking the child to select 3, 7, and 15 items (like beans or stones) from a pile, basic addition and subtraction, and word problems. Several CLAs incorporate culturally appropriate

⁴ Tools measuring early childhood and early primary grade skills

math problems. For instance, Uwezo (‘capability’ in Kiswahili) Kenya includes a section for ethno-math, adding prices (in Kenyan shillings). of everyday objects, like mangos and milking jelly.

Table 4. Types of assessment tools by numeracy domain

Category	Tools	Math Proficiency (e.g., problem-solving, reasoning)	Number Knowledge (including operations)	Measurement	Statistics and Probability	Geometry	Algebra
System-monitoring tools							
International assessments	PISA-D, TIMSS, PIRLS, LaNA	Lightest blue	Darkest blue	Lightest blue	Darkest blue	Darkest blue	Lightest blue
Regional assessments	LLECE, SEACMEQ, PASEC, PILNA, SEA-PLM, UNRWA MLA	Darkest blue	Darkest blue	Darkest blue	Lightest blue	Lightest blue	Lightest blue
Household	MICS-ECDI, MICS-FLS	No color	Darkest blue	No color	No color	No color	Lightest blue
Multi-purpose tools							
Foundational skills	MELQO, EGMA, Numeracy Boost	Lightest blue	Darkest blue	Lightest blue	No color	Lightest blue	Lightest blue
Citizen led	ASER, UWEZO, LearnNigeria, etc.	Lightest blue	Darkest blue	No color	No color	No color	No color
Designed for EiE contexts	HALDO, IDELA	Lightest blue	Darkest blue	Lightest blue	No color	Lightest blue	No color

-  Darkest blue indicates that all tools in the category include sub-domain
-  Lighter blue indicates that half or more tools in the category include sub-domain
-  Lightest blue indicates that one or a few tools in the category include sub-domain
-  No color indicates that no tools in the category include sub-domain

Summary

Quality education is a key component of humanitarian response in conflict- or crisis-affected countries. Measuring learning in such contexts can be challenging. Assessments used in other contexts can be too difficult for children living in these circumstances, not contextually valid, or fail to capture the contextual factors that influence learning outcomes.

In general, we found that guidance documents used in EiE settings, including the SDGs, did not provide robust details on the academic domains that should be measured or specifics on how learning assessments should be contextualized. Further work has been conducted by organizations such as the UIS

to operationalize some of the global frameworks and link them to existing assessment tools. In other cases, the guidance documents remain open to interpretation and, as a result, they are not used extensively in the monitoring and evaluation of learning outcomes at a national or program level.

Of the 30 measurement tools examined, only 3 were developed specifically for EiE contexts: HALDO, the MLA, and OLA. Among the remaining tools, we found that

- the majority of learning assessment studies do not gather data on refugee or IDP status;
- many learning assessment studies do not construct a sample to ensure the inclusion of refugees and IDPs or a sample to use in conflict- or crisis-affected regions—in other words, a nationally representative sample does not occur;
- most learning assessment studies are conducted in schools, some in homes, which means that children affected by conflict and crisis and those who are out of school are unlikely to be included in a study unless they have been resettled in a home and enrolled in a government school;
- even children who are enrolled in government schools are sometimes excluded from assessments, such as EGRA, SEACMEQ and PASEC;
- the protocols for most large-scale assessments require testing in the national language and, for the safety of the test administrators, exclude conflict zones.

All the tools examined have conducted some type of reliability or validity testing, but the degree to which these results are transparent varies. We found that, among the constructs measured, decoding, reading comprehension, and number sense were widely captured by most tools. This finding was also reflected in a small sample of national assessments we looked at.

The guidance documents, measurement tools, and program approaches mapped in this paper have the potential to increase equity for children in conflict-affected and fragile contexts, but by and large they do not currently do so. Governments and non-government actors can include EiE settings in the measurement of learning outcomes by including EiE contexts in the development and validation of measurement tools, and by sampling areas of a country affected by conflict or crisis. Collecting contextual data and using that information to explain differences among children is another way researchers and policy-makers can improve equity in EiE settings. Finally, it is important that the data gathered by these measurement efforts are used to start a conversation about the quality of learning opportunities for children in conflict- or crisis-affected areas, and that the results are acted on.

Part 1. Background

The global landscape

The Sustainable Development Goals (SDGs) ushered in an ambitious agenda for the education sector, in particular SDG4, which aims to “ensure inclusive and equitable quality education and promote lifelong learning opportunities for all” by 2030. These goals are the culmination of more than a decade of reform efforts that have shifted gradually from a focus on school access and attendance to improving the quality and equity of the education provided. A critical aspect of realizing this shift is the increased use and availability of learning assessments. Recent flagship reports, such as UNESCO’s annual Global Education Monitoring Report and the World Bank’s 2018 World Development Report—the first in this series to focus on education since its start in 1978—urged making investments in learning as well as schooling.

More and better assessment data will be required to develop and implement effective learning programs and policies. The number of regions, countries, and education interventions that collect learning assessment data continues to increase, which has sparked global awareness and spurred efforts to improve the quality of education. However, in education in emergency (EiE) settings, data on learning remains sparse.

The sector-wide shift in focus toward equity and learning requires a common understanding and alignment of measurement tools between EiE actors and the implementers and funders working in stable development contexts, and the host governments. Recognizing this, the Inter-agency Network for Education in Emergencies (INEE) Education Policy Working Group (EPWG) initiated a Quality and Equitable Learning Outcomes (QELO) work stream to “create a shared understanding of equitable quality learning outcomes and measurement of quality learning outcomes across humanitarian and development education programs” (Nirrengarten, Pulfer, & Weisenhorn, 2018, p. 1). To ensure that the QELO work builds on existing guidance documents and measurement tools, the group commissioned two separate mapping exercises of academic and social and emotional learning (SEL) program approaches,⁵ measurement tools, and monitoring and results frameworks, their aim being to map what quality education in emergencies constitutes at the program and individual learning levels, to inform the linkages with global measurements.

This report contributes to the academic component of the QELO work stream by outlining the landscape of guidance documents, tools, and program approaches used to measure learning outcomes in EiE contexts and providing an analysis of the links and alignment among them in order to inform a more streamlined measurement of learning outcomes in emergency contexts.

⁵ The SEL mapping was conducted by the EASEL Lab at the Harvard Graduate School of Education.

INEE's work to date

INEE was conceptualized in 2000, during the strategy session on EiE at the World Education Forum in Dakar. INEE aims to build on and consolidate existing networks and efforts by UN organizations, the World Bank, bilateral donors, and the more than 20 NGOs engaged in EiE programming. INEE serves its members through community-building, by convening diverse stakeholders in thematic task teams and working groups, knowledge management, advocating and amplifying ideas, training and capacity development, facilitating collective action, and providing members with resources and support.

To support the measurement of learning outcomes in emergency settings, INEE has developed several iterations of the *INEE Minimum Standards Handbook*, which contains 19 standards and accompanying key actions to enhance the quality of educational preparedness, response, and recovery; increase access to safe and relevant learning opportunities; and ensure accountability. The standards in the handbook are organized into five domains: foundational standards, access and learning environment, teaching and learning, teachers and other personnel, and education policy formulation and implementation. An accompanying EiE toolkit contains practical, field-friendly tools to guide EiE implementers on each of the INEE standards.

INEE also has worked on social and emotional learning (SEL) and psychosocial support (PSS) in EiE through guidance notes and roundtables. The INEE PSS-SEL Measurement Reference Group works on tools for measuring SEL in crisis contexts. INEE also has developed tools and guidance documents for EiE teachers on pedagogy for learning, adapting curricula, teacher training, and professional development, and for safer school construction, teacher compensation, and education finance.

Considering INEE's previous work and comparative advantage in the sector, it is well placed to promote alignment around measurement of comprehensive learning outcomes among the broad spectrum of actors it represents, including donors, implementers, and researchers. Moreover, its representation in meaningful EiE bodies will help advance policies. INEE also has the convening power to drive a shared agenda around core EiE issues, and in particular, ensuring that work on these issues is grounded in a strong knowledge and evidence base.

Assessing learning outcomes in EiE contexts

Providing quality education in conflict- or crisis-affected areas is imperative to keep children safe and help them rebuild their lives, but measuring learning in such contexts can be challenging. Furthermore, existing assessments do not represent the breadth of skills that are important for EiE, including those that equip children to handle the challenges of an increasingly complex life, in the emergency setting and afterward. Social and emotional skills, critical thinking, and life and coping skills are often taught in EiE contexts, but they may not be captured through formal measurements.

Fragmentation is acute between actors in the EiE sector, and between the EiE sector and more stable development sectors, in terms of how to implement common learning standards and assessments. Although several donors and implementers, including USAID, Porticus, the International Rescue Committee (IRC), New York University (NYU), and UNICEF Middle East, are working on this issue, the various initiatives work in isolation from each other and separately from the larger initiatives geared toward measuring SDG4.⁶

The humanitarian sector faces a number of unique challenges in delivering and measuring learning in emergencies. Moreover, without a consistent set of goals as to what children should know and be able to do, it is difficult to build a functional learning assessment system. Implementing actors in crisis contexts often use program-specific measurement systems, as existing assessment systems often are not set up to capture populations on the move, or they respond inadequately to additional strain. This results in a fragmented system with non-comparable data and few or no baseline reference points. In addition, when refugee or IDP populations are integrated into national education systems, their status often is not appropriately labeled. While this may be a positive step in terms of reducing stigma, it makes tracking progress impossible and makes it difficult for policy-makers to address the children's specific needs.

National governments, particularly those of countries affected by conflict and crisis, are often unable to collect data and manage assessment systems, and it is unclear which actors would fill this mandate in their place. The nature of short-term funding for education and learning projects in crisis contexts limits the implementing agencies' ability to help strengthen national and local entities' capacity to collect data. This fragmentation can be amplified in situations where parallel education systems exist—for example, in a context where some refugees attend the public schools in their host community while others reside in camps where education services are provided by the international community or local NGOs.

Many traditional national and international assessments are administered only in the classroom, but a large proportion of children in emergency settings are not in school. Only 50% of refugee children have access to primary school, compared to 90% of children globally, and only 22% of refugee youth are in secondary school compared to 84% globally. Therefore, measuring learning outcomes for all children both in and out of school requires alternative strategies such as household-based surveys, which are costlier and require modified administration approaches in conflict- and crisis-affected settings. While it might be easier to conduct assessments in a tightly controlled camp environment, parents in that environment could be unwilling to allow their children to be assessed. In some humanitarian situations, such as the Rohingya refugee crisis in Bangladesh, neither the host country nor the country of origin allows their national curriculum to be used with the refugee children. (McPherson & Paul, 2019)

Finally, children and youth who have faced conflict and crisis may bear a cognitive load far greater than that of their peers. This means that conflict- or crisis-affected populations would be at a disadvantage on

⁶ One effort is the Middle East and North Africa (MENA)-based practitioner-research consortium to develop and validate measurement tools for EiE. More recently, UNICEF has developed a framework for learning outcomes for the Middle East that includes life skills.

standardized tests that have been normed on populations that have not faced the same difficulties. Traditional test protocols also may not provide support and protection in keeping with the principle of “do no harm,” as they may cause frustration and anxiety in test-takers.

Methodology

Numerous regional and international assessments are used to measure numeracy and literacy, each with different purposes, target populations, methodologies, procedures, and relationships to guidance documents. In this paper, we categorize and evaluate these assessments in three parts:

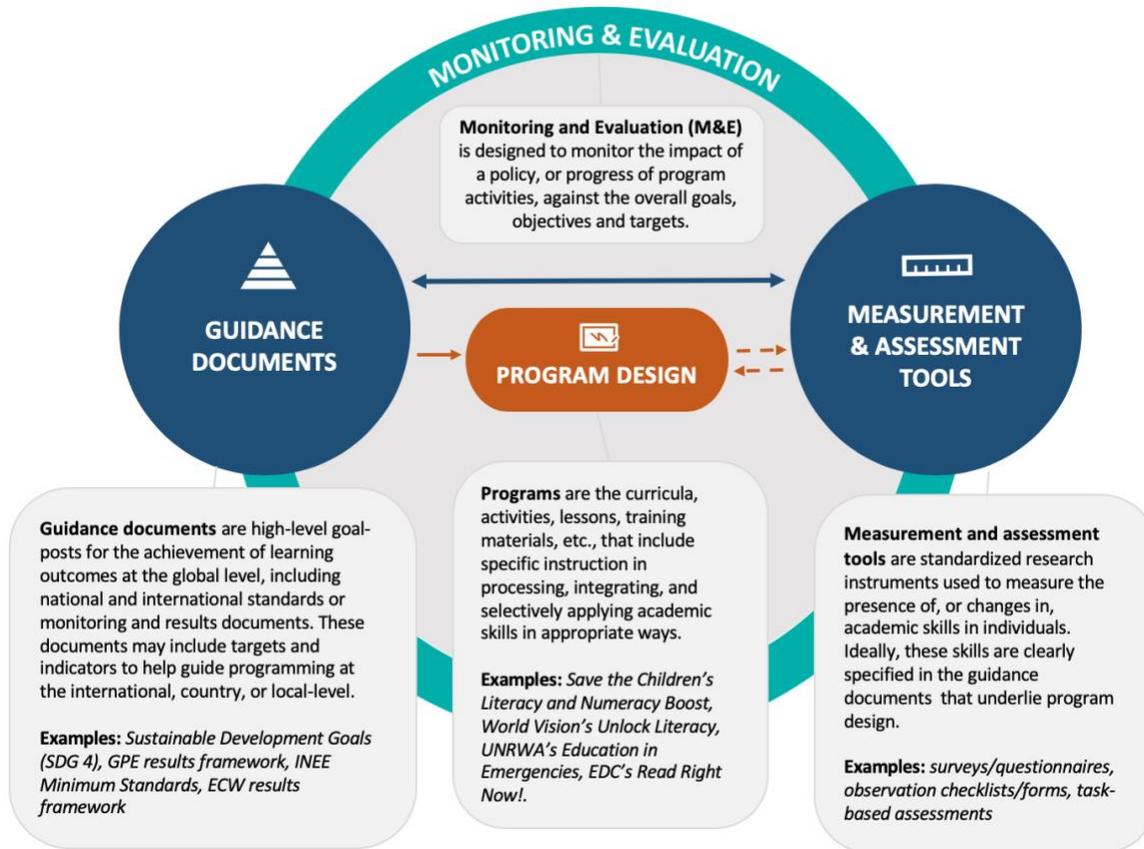
- Part 1 compares the tools according to each one’s purpose, design, target population, and various other parameters.
- Part 2 illustrates commonalities and differences across the domains and sub-domains each tool assesses.
- Part 3 summarizes linkages among the measurement tools and frameworks and their application in EiE contexts.

Parameters and taxonomy

The purpose of the INEE QELO Mapping Project is to identify and map the assessment tools and guidance documents currently being used in the EiE sector. We adopt the following heuristic to show the connections between programs, measurement and assessment tools, guidance documents, and program monitoring and evaluation systems (see Figure 3).⁷

⁷ Because work by the EASEL Lab SEL team was underway several months before the academic team was in place, the parameters and taxonomy were adopted and adapted from the methodology used for the SEL component in order to maintain alignment between the two mappings.

Figure 3. Relationship between guidance documents, program design, measurement/assessment tools, and monitoring and evaluation.



Guidance documents are high-level monitoring and results frameworks. They often are written policy documents or standards that provide guidance and goalposts for the achievement of learning outcomes at the global level. Guidance documents may be published by multilateral organizations, national governments, and/or influential NGOs in the EiE sector.

Measurement and assessment tools are standardized research instruments used to measure the presence of, or changes in, individuals' academic skills (e.g., literacy, numeracy). They may capture skills directly or measure other indicators as a proxy for a particular skill or characteristic (e.g., mother tongue, language of instruction). They can be administered with paper-and-pencil tests, by asking a child to perform a task and recording their response, or on a computer or tablet.

Measurement and assessment tools are sometimes categorized as assessments *of* learning, which capture what an individual has already learned, and assessments *for* or *as* learning, which capture information in real time that can be used to modify instruction to improve learning. An assessment *of* learning is sometimes called summative assessment, which includes sample-based studies of student learning and

examinations. An assessment *for* or *as* learning is sometimes called formative assessment, which includes classroom-based assessments, as well as the informal ways teachers assess students' abilities and how they respond to that information (Clarke, 2012). The tools considered in this review are primarily concerned with assessments *of* learning, with some attention to assessments *for* learning as they relate to programmatic approaches. The latter is more likely to be done at the individual program level and to be contextualized, and it is less likely to be done in a systematic, documented manner, which makes it difficult to cover comprehensively in this paper.

Influential programmatic approaches are developed by international organizations to shape their programmatic work across countries. These approaches provide case examples of the connection and divergence between what is in the guidance documents, program designs, and assessment tools of a particular approach or intervention.

As shown in Figure 3, the relationships between guidance documents, measurement tools, program approaches, and monitoring and evaluation are multifaceted.

1. Guidance documents and measurement/assessment tools inform one another. Guidance documents indicate which learning outcome constructs are critical/relevant and which tools should be used to measure these constructs.
2. Measurement/assessment tools generate the data needed to reflect on and revise the skills represented in the guidance documents that are based on data.
3. Guidance documents inform program design, and inversely programs rely on international, country, or local guidance documents to structure the programmatic curricula so that it meets the high-level learning outcomes specified.
4. Program design influences measurement/assessment tools, and vice versa, as measurement tools are used to assess the impact of programming on instruction and the outcomes targeted by programs. Measurement/assessment tools are in turn often adapted for use by various programs. Once outcomes are measured, program designers use data to revise and adapt their strategies.
5. Monitoring and evaluation cuts across these categories. Monitoring and evaluation systems and frameworks often include measurement/assessment tools that are used at the local, national, and international levels for both formative and summative evaluation. Similarly, guidance documents and programmatic approaches both shape monitoring and evaluation, as they inform the selection of outcomes, targets, and indicators that will be monitored and evaluated.

Selection criteria

This analysis reviewed 31 measurement tools (23 distinct approaches, with a group of 8 citizen-led assessments (CLAs) comprising the remaining tools), 8 guidance documents, and program approaches from 4 organizations. Three key factors determined their selection: coverage, quality, and relevance. We located tools and frameworks through desk research, conversations with key stakeholders, and an online survey of INEE members. The full list of stakeholders consulted appears in Annex A.

The tools included cover children and youth ages 3 to 18 in conflict-affected contexts around the globe, including Africa (Francophone, Anglophone, Lusophone), South Asia, Latin America, the Pacific Islands, Eastern Europe, Central Asia, and the Middle East (see Figure 4). A full list of the crisis-affected countries covered by each tool and framework is available in Annexes B and C.

Guidance documents and program approaches were included based on the following criteria:

1. They are influential in setting the agenda and/or guiding the field of education globally or regionally, according to key stakeholders.
2. Our team was able to access and acquire their materials, either on our own or from the organization or program staff.
3. They address literacy and numeracy learning outcomes.

Measurement tools were included based on the following criteria:⁸

1. They address literacy and/or numeracy learning outcomes.
2. They are used in at least one country where EiE is provided.⁹ We identified 61 countries for this exercise by combining the UNHCR Refugee Situations list,¹⁰ the UN Office for the Coordination of Humanitarian Affairs (OCHA) Humanitarian Operations list,¹¹ and suggestions from INEE QELO members (see Figure 4). The tools selected can be administered by local or regional programs or organizations, international or multilateral programs or organizations, or national governments and education systems.
3. They assess education levels from pre-primary through lower-secondary education, defined as levels 0-2 by the International Standard Classification of Education (ISCED), which are the education levels most education systems consider compulsory. (UIS, 2012) Descriptions of the ISCED levels are found in Table 5. Education levels covered in the mapping exercise.
4. They have been used in at least one full-cycle assessment process (design, collection, analysis, and dissemination).
5. They have materials that our team was able to access and acquire, either on our own or through organization or program staff members, which include the actual tool or a framework describing the types of items in the tool.
6. They are codable materials (i.e., they explicitly define or describe individual skills, competencies, behaviors, strengths, etc.). Measurement tools or their corresponding frameworks must clearly define or describe the literacy and numeracy skills they measure, and/or we must be able to code the actual task or observation or questionnaire items.

⁸ Several tools initially considered were excluded because they did not fit the criteria, such as International Civic and Citizenship Education Study (ICCS), which does not measure literacy and numeracy. Others were excluded because they were not administered in EiE countries, such as the recently developed Early Grade Writing Assessment (EGWA) sponsored by UNESCO.

⁹ INEE defines “education in emergencies” as including conflicts, situations of violence, forced displacement, disasters, and public health emergencies. Education in emergencies is a wider concept than emergency education response, which is an essential part of it. See <https://archive.ineesite.org/en/education-in-emergencies>.

¹⁰ See <https://data2.unhcr.org/en/countries/>.

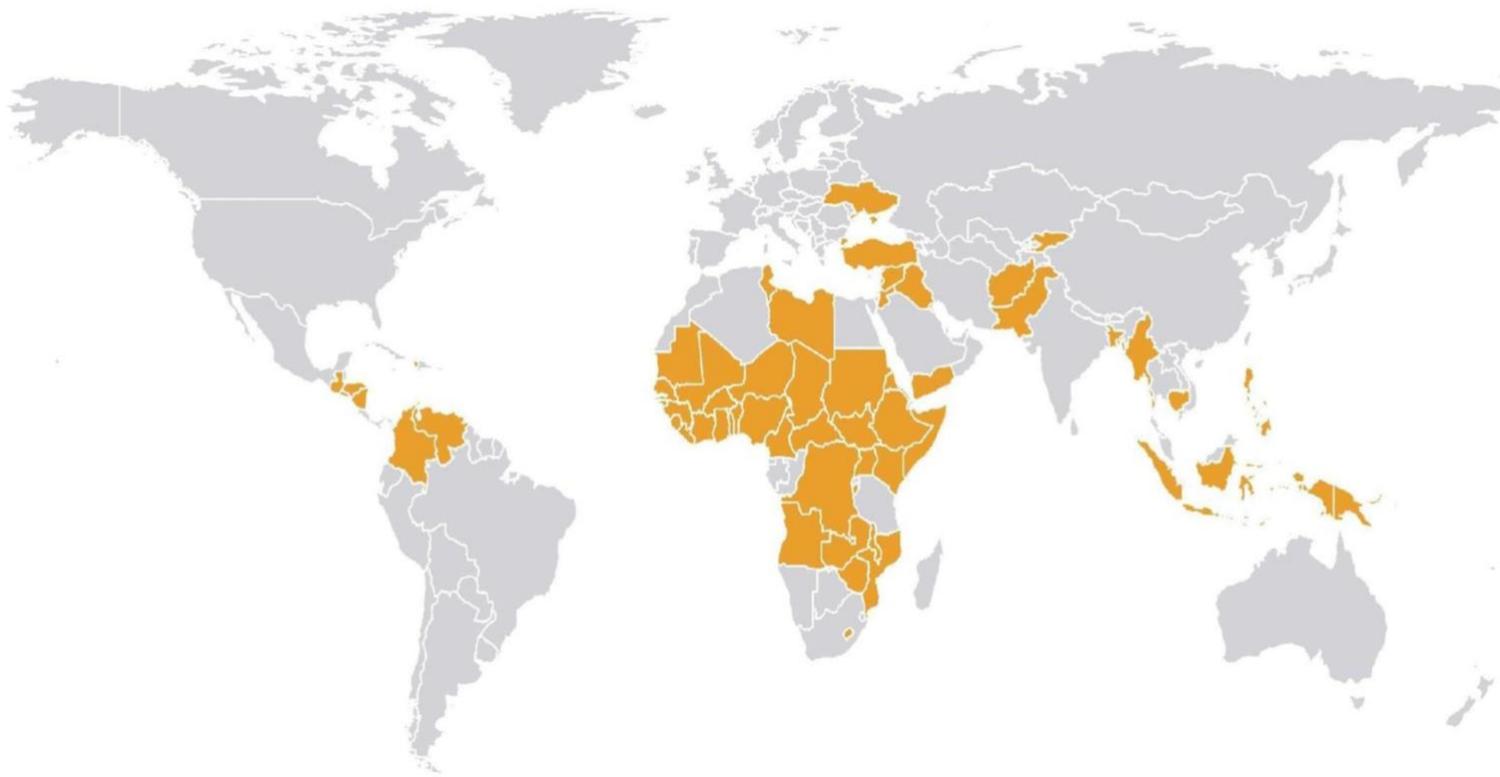
¹¹ See <https://www.humanitarianresponse.info/en/operations>.

Establishing reliability and validity was not a criterion for inclusion, but all tools that met our criteria claimed to have established reliability and validity. However, they did not all have accessible information about their process of establishing reliability and validity.

Table 5. Education levels covered in the mapping exercise

ISCED Level	Name	Description
0	Pre-primary education	The initial stage of organized instruction designed to introduce very young children to a school-type environment and to develop their cognitive, physical, social, and emotional skills. Designed for children from age 3 to the start of primary education.
1	Primary education or first stage of basic education	Normally starting between the ages of 5-7, designed to give a sound basic education in reading, writing, and mathematics, along with an elementary understanding of other subjects.
2	Lower secondary education or second stage of basic education	Designed to complete basic education, usually on a more subject-oriented pattern. Builds on the learning outcomes from primary education (ISCED level 1) and aims to lay the foundation for lifelong learning and human development.

Figure 4. Countries of interest for analysis of measurement tools



Conflict or crisis-affected countries

Afghanistan	Congo, DRC	Honduras	Mali	Senegal	Uganda
Angola	Cote d'Ivoire	Indonesia	Mauritania	Sierra Leone	Ukraine
Bangladesh	El Salvador	Iraq	Mozambique	Solomon Islands	Venezuela
Benin	Eritrea	Jordan	Myanmar	Somalia	Yemen
Burkina Faso	Ethiopia	Kenya	Nicaragua	Somalia	Zambia
Burundi	The Gambia	Kyrgyzstan	Niger	South Sudan	Zimbabwe
Cambodia	Ghana	Lebanon	Nigeria	Sudan	
Cameroon	Guatemala	Lesotho	Pakistan	Syria	
Chad	Guinea	Liberia	Palestine	Togo	
Central African Republic	Guinea-Bissau	Libya	Papua New Guinea	Tonga	
Colombia	Haiti	Malawi	Philippines	Tunisia	
				Turkey	

Part 2. Overview of Tools, Frameworks, and Program Approaches

Guidance documents

Guidance documents provide high-level guidance for the monitoring of learning outcomes and program design for all levels of education. They often are written policy documents or standards frameworks. The review includes two global frameworks, three regional frameworks, one minimum standards document for EiE settings, and two results frameworks for specific organizations (Global Partnership for Education and Education Cannot Wait). See Annex B for a description of all of the guidance documents reviewed.

Although these documents mention reading and mathematics as core competencies, none includes more specific constructs, domains, or sub-domains of literacy and numeracy. The SDGs, as agreed in 2015, do not provide common definitions or constructs of minimum proficiencies for what should be measured in reading and mathematics. In August 2019, the Global Alliance to Monitor Learning (GAML) met to agree on minimum proficiency standards for reading and numeracy (USAID & UIS, 2019). Table 6 and

Table 7 present the recently developed global minimum proficiency levels.

Table 6. Reading minimum proficiency levels—UNESCO Institute for Statistics

Grade 2	Learners read aloud some common words and comprehend most directly stated information in a short, simple text. They make simple inferences when a longer text is read aloud to them.
Grade 3	Learners read texts fluently, identify the topic of a text, locate directly stated information, and make very simple inferences from short written texts. Learners understand explicit details and make simple inferences based on directly stated clues when a text is read aloud to them.
Grade 4	Learners read aloud with expression. They find information in grade-level texts and use word knowledge and prior experiences to interpret and make reflections.
Grade 5	Learners infer the meaning of most unknown words and expressions by using clues in the words or sentences. They locate prominent information in texts, recognize key ideas, infer points of view and causal relationships, and connect ideas with their personal knowledge and experience.
Grade 6	Learners locate explicit details, differentiate fact from opinion, recognize the purpose of a text, identify the main and prominent secondary ideas, relate them to their personal knowledge and experience, and draw basic conclusions.

Table 7. Mathematics minimum proficiency levels—UNESCO Institute for Statistics

Grade 2	Learners demonstrate skills in number sense and in computation involving whole numbers up to 100, reading simple data displays, shape recognition, and spatial orientation.
Grade 3	Learners demonstrate skills in number sense with numbers up to 1,000, with computations within 100, in identifying and representing common fractions, in measuring and comparing lengths and weights, in recognizing shapes and figures regardless of their orientation, in interpreting simple categorical data displays, and recognizing simple, increasing numerical patterns.
Grade 4	Learners demonstrate skills in number sense with numbers up to 10,000, with computations within 1,000, with real-world problems involving 4 operations, in measuring and comparing lengths, weights, and capacities/volumes, in recognizing 2D shapes by their attributes, and reading and interpreting simple categorical data displays.
Grade 5	Learners demonstrate skills in number sense with numbers up to 100,000 and decimals to hundredths place, in multiplication of two-digit numbers, and addition and subtraction of fractions with the same denominator, in real-world problems involving perimeter, in recognizing three-dimensional figures by their attributes, in organizing categorical data into simple displays.
Grade 6	Learners demonstrate skills in number sense with decimals to thousandths place, in the addition and subtraction of fractions and decimals, in converting between simple fractions, decimals, and percentages, in solving problems involving elapsed time, area, and proportional reasoning, in identifying different views of three-dimensional figures, in interpreting complex, categorical data displays, and in identifying position, direction, and coordinates on maps and graphs.

The SDG4 target and indicator framework does list “conflict-affected” as a dimension of equity that should be captured as data become available. The indicators for SDG4.5.1 are parity indices (female/male, rural/urban, bottom/top wealth quintile, and others such as disability status, indigenous peoples, and conflict affected, as data become available) for all education indicators on this list that can be disaggregated.

Several guidance documents, including Reimagining Life Skills and Citizenship Education in the Middle East and North Africa: Conceptual and Programmatic Framework and the INEE Minimum Standards, also highlight what kinds of learning assessments to use and provide general direction on how to conduct learning assessments. Others only discuss learning measurement generally, and some of these, like the Pacific Regional Education Framework, explicitly state that there are cultural reasons for omitting specific guidance, as context influences the types of skills that are taught and measured.

As the agreed-to global reference point, the SDGs and related indicators are widely influential, and many regional and organization-specific learning frameworks align with SDG guidance. Regional frameworks often define themselves as a regional extension of the SDGs, essentially owning the goals and adapting them to regional aspirations. For example, the Continental Strategy for Education in Africa uses the SDG indicators to monitor learning progress. Others, like the Pacific Regional Education Framework, do not specify targets but state generally that the region will maintain its alignment with SDG4 and prioritize a coordinated approach to SDG-related monitoring. In line with the equity goals of SDG4, all frameworks highlight the importance of disaggregating data by gender, socioeconomic status, disability, language, context, and other indicators of marginalization. However, disaggregating by refugee or IDP status was not included in the global guidance documents reviewed.

In the specific area of education in emergencies, the INEE Minimum Standards are the sole reference point. The standards are referenced in several of the regional frameworks, including Reimagining Life Skills and Citizenship Education in the Middle East and North Africa. The ECW Results Framework also uses the INEE Minimum Standards as part of the benchmark for instructional/learning processes across ECW-supported programs. The INEE Minimum Standards do not recommend, require, or align with specific assessment tools. The Global Partnership for Education (GPE)'s framework monitors the percentage of its developing country partners that improved their learning outcomes in basic education, and whether they have a learning assessment system that meets quality standards. It does not, however, specify the domains or tools that should be used.

National standards and curricula are also important guidance documents for countries when it comes to providing guidance for measuring learning. Experts agree that, while a common core of learning areas lends itself to global and regional alignment and comparison, many aspects of learning should not necessarily be comparable across countries. The vast majority of countries have defined learning objectives through their official curriculum and associated national assessments. When discussing regional or international assessments or minimum levels of competency, countries may be concerned that their curriculum is not sufficiently represented in the assessment or the standard. This can influence the way cross-national assessments are used and perceived.

Assessment tools

The 30 assessment tools described here are cross-national, meaning that they are used in more than one country. The tools include NGO assessments, regional and international assessments, citizen-led assessments, and assessments developed by international agencies. Individual national assessment tools were not included in the mapping for several reasons. First, all of the regional and international assessments and some others were constructed by reviewing curricula and standards from countries participating in the assessments. Given this, we can make some assumptions about the content of the national assessments for countries participating in regional and international assessments. We also reviewed the IBE/UIS studies, "Monitoring progress towards SDG 4.1: Initial analysis of national assessment frameworks for reading" (2017) and "Monitoring progress towards SDG 4.1: Initial analysis of national assessment frameworks for math" (2018), which mapped national assessments across 53 countries. We felt this was a sufficient number and did not see it as a good use of resources to replicate it or review additional countries. Second, most national assessment tools did not meet our criterion for accessibility, and having only a few conveniently sampled national assessments along with the 30 cross-national assessments would have made analysis difficult. A brief description of the content of several national assessments appears in Part 2.

Validity and reliability

The MELQO overview provides a concise explanation of the types of reliability and validity possible for learning assessment tools (see Table 8).

Table 8. Types of psychometric properties¹²

Psychometric Property	Definition
Construct validity	Items accurately index key constructs
Concurrent validity	Items correlate well with other established sets of items
Inter-rater reliability	Items can be reliably collected and do not vary from one assessor to the next
Internal consistency	Items within a scale measure the same underlying construct
Cross-cultural relevance	Items are relevant across contexts
Predictive validity	Items reliably predict children’s development and future learning

All measurement tools assessed in this study have been designed and piloted to determine some psychometric standards of validity and reliability, although the transparency of the results and the extent to which validity and reliability have been established varies (see

¹² UNESCO, UNICEF, Brookings Institution, & World Bank, 2017

Table 9). The tools are typically piloted and validated on an as-needed basis as part of the adaptation process when used in a new context or language.

For tools that are to be used in various contexts and across multiple languages, such as EGRA and EGMA, detailed toolkits provide instruction on how to adapt the instrument. Three tools—Save the Children’s Holistic Assessment of Learning and Development (HALDO), the UNRWA Monitoring Learning Achievement (MLA), and EDC’s Out-of-School-Youth Literacy Assessment (OLA)—have been validated in an EiE setting. For instance, HALDO was piloted on 852 children from 27 centers and schools in the Dadaab refugee complex in Kenya (Krupar, D’sa, Westrope, & Johna, 2019). The study found strong evidence of reliability and validity in identifying a baseline of learning for 4- to 12-year-olds, although the tool still requires contextualization to local social and cultural norms when adapted for other locations.

Table 9 describes the tools, the populations the tools were developed for and piloted with, and a brief summary of reliability and validity.

Table 9. Characteristics of the populations, reliability and validity of assessment tools

Assessment	What population was it developed for?	Who was it piloted with?	Who validated the tools?	Reliability/Validity Information
Citizen-Led Assessments (includes 8 CLAs reviewed for this paper)*	Students and out-of-school children in developing countries	1,207 children in Kenya	NGOs in each country	The results of a concurrent validity study conducted by the Australian Council for Educational Research revealed that Uwezo tools are of similar difficulty or easier than core EGRA/EGMA tools, but the correlation between the tests is high, indicating that they are measuring the same or very similar constructs. Reliabilities of the Uwezo assessments were generally lower than those of core EGRA/EGMA, indicating that Uwezo can explain a lower percentage of the variation in children’s performance.
Early Grade Mathematics Assessment (EGMA)	Students in developing countries	5,392 grade 1 and grade 2 children in two developing countries (countries unnamed in EGMA toolkit)	RTI, and each organization that administers it in a new country/language	The authors evaluated validity of the Core EGMA based on content, response processes, internal structure, linkage to other variables, and consequences of testing. They generally found that EGMA is trustworthy and meaningful for making decisions to support student learning in early grade mathematics.
Early Grade Reading Assessment (EGRA)	Students in developing countries	Varies in each country where administered	RTI, and each organization that administers it in a new country/language	Each EGRA must be validated for the language and country in which it is being used. Content validity is established through workshops with stakeholders in each country. Reliability should be determined at a minimum through establishing Cronbach’s alpha and validity through analysis of internal consistency; according to the Early Grade Reading Barometer, this has been done for each of the language pilots.
Holistic Assessment of Learning and Development Outcomes (HALDO)	Children in emergency settings	852 children ages 4-12 in Dadaab, Kenya	Save the Children	Authors consider reliability and validity strong. Scores increased predictably by age, signifying the tool measures skills that follow expected child development trajectories.
International Development and Early Learning Assessment (IDELA)	Young children in developing countries, many of them in emergency settings	Piloted across 12 countries: Bangladesh, Bhutan, Egypt, Ethiopia, India, Indonesia, Mali, Malawi, Mozambique, Pakistan, Rwanda, and Zambia	Save the Children	Construct validity testing used exploratory and confirmatory factor analyses and identified that the tool contained four distinct factors aligned with the hypothesized domains, as well as a single over-arching construct of children development (Wolf et al., 2017).

Assessment	What population was it developed for?	Who was it piloted with?	Who validated the tools?	Reliability/Validity Information
Literacy and Numeracy Assessment (LaNA)	Students in developing countries	Piloted in Haiti, Macedonia, Nigeria, Pakistan (Punjab), and Serbia	The TIMSS & PIRLS International Study Center at Boston College	Validation is ongoing. Early results indicates the difficulty level of the assessment is a good fit in Punjab and Haiti, while in other countries further calibration is needed.
Literacy Boost	Students in developing countries	Varies in each country where administered	Save the Children	Individual reports for each country, generally high levels of inter-rater reliability and good internal consistency
Measuring Early Learning and Quality Outcomes (MELQO) Measure of Development and Early Learning (MODEL)	Young children in developing countries	1,142 children in Lao PDR, Mozambique, Madagascar, and Tanzania	Various researchers: ECD Measure Group, World Bank, NYU, and others	Concurrent validity established through comparing parent/teacher reports and direct child assessments. Adequate inter-rater reliability and internal consistency. Cross-cultural relevance established for some items.
Multiple Indicator Cluster Survey (MICS-) Early Childhood Development Index (ECDI)	Young children in developing countries	1,102 children in Jordan and 1,004 children in the Philippines	UNICEF	Content validity of the items was confirmed in both countries with acceptable test-retest and inter-rater reliabilities and acceptable internal consistency.
MICS- Foundational Learning Skills (FLS)	Students and out-of-school children in developing countries	130 children in four villages in Kenya	UNICEF	Strong inter-rater reliability, demonstrated internal consistency, and most tasks consistent with EGRA/EGMA tasks
Numeracy Boost	Students in developing countries	Varies in each country where administered	Save the Children	Individual reports for each country, generally high levels of inter-rater reliability and good internal consistency
Out-of-School Youth Literacy Assessment (OLA)	Participants in non-formal education programs	2,000 participants in a youth literacy program in Liberia	Education Development Center	OLA's component subtests have strong predictive relationships to the development of reading ability in alphabetic languages.
Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC)	Students in francophone African countries	Varies in each country where administered	PASEC team at CONFEMEN, with national teams	Internal consistency was calculated for each index and for each country. According to PASEC procedures, the minimum internal consistency threshold (Cronbach's alpha) as part of the contextual data analysis is 0.60 and the majority of countries have reached this.
Pacific Islands Literacy and Numeracy Assessment (PILNA)	Students in Pacific Islands countries	Validated in each country where administered	Educational Quality and Assessment Division (EQAP): Pacific Community	Concurrent validity established through comparing literacy and numeracy outcomes against other national sources (e.g., NGO surveys and research, national census)

Assessment	What population was it developed for?	Who was it piloted with?	Who validated the tools?	Reliability/Validity Information
Progress in International Reading Literacy Study (PIRLS)	Students in participating countries—began with a small number of European and North American countries but has expanded to all regions	Pilots with 200 students in each country for each administration	National teams, with support from IEA and the TIMSS & PIRLS International Study Center at Boston College	Reliability and cross-country validity generally very high.
Programme for International Student Assessment for Development (PISA-D)	Grade 7 students and out-of-school 14- to 16-year-olds in Bhutan, Cambodia, Ecuador, Guatemala, Honduras, Paraguay, Senegal, and Zambia	Field-trialed with 22,250 students in 7 countries, and then piloted with nearly 37,000 children in 8 countries	OECD, Educational Testing Services (ETS), participating countries, development partners, and technical partners	Construct validity established early in the project. Most items were already used in PISA 2015. The main survey data underwent a multi-step analysis process, including evaluating the data quality, item analysis, and IRT scaling, in order to enable placing results on the main PISA scale.
Southeast Asia Primary Learning Metric (SEA-PLM)	Students in East Asia and Pacific countries	Varies in each country where administered	National teams, with support from SEA-PLM and ACER	Information on reliability and validity not provided, other than it occurs at the country level.
Southern and Eastern Africa Consortium for Monitoring Educational Quality (SEACMEQ)	Students in anglophone African countries	First piloted in a small number of primary schools in South Africa, then in each country where administered	SEACMEQ, with national teams	Statistical and content validity and reliability checks on pilot data were carried out by the SEACMEQ Coordinating Centre, which declared them ready for the field (no additional info).
School-based Test About Reading (STAR)	Students in developing countries	Not found	World Vision	Tool is based on EGRA, Literacy Boost, and a previous World Vision tool, so content validity is assumed through that process.
Third Regional Comparative and Explanatory Study (TERCE)	Students in Latin American countries	31,324 students across 15 Latin American countries	UNESCO LLECE	During the design and construction of the exam, there were several rounds of revision. Items were maintained, modified, or discarded based on validity, whether each item effectively evaluated what it was designed to evaluate. They were also revised based on alignment to the test and to the context, clarity, and correctness. Each test item was evaluated using two methods: the Theory of Response Items using the Rasch model, as well as Psychometric Analysis using the Classical Theory of the Test by block and by booklet.

Assessment	What population was it developed for?	Who was it piloted with?	Who validated the tools?	Reliability/Validity Information
Trends in Mathematics and Science Study (TIMSS)	Students in participating countries—began with a small number of European and North American countries but has expanded to all regions	Pilots with 200 students in each country for each administration	National teams, with support from IEA and the TIMSS & PIRLS International Study Center at Boston College	Reliability and cross-country validity generally very high.
UNRWA MLA	Palestinian refugee students	Pilots in four schools each in Gaza, Jordan, Lebanon, and the West Bank	UNRWA	Validity and reliability results not specified. Results of pilots were used to select items for final version.

Inclusion of children in emergency contexts

Two assessment tools, HALDO and the UNRWA MLA, were developed specifically for EiE contexts, and IDELA was created for developing countries more generally, but was also piloted in several EiE countries. Some tools were designed for general populations in developing countries but are being used in refugee camps. In Uganda, in the four districts with the highest concentration of refugees, Uwezo conducted a study that compared the basic literacy and numeracy of refugee children to that of non-refugee children. They found that refugee children outperformed their counterparts in the 120 villages and 2,186 households included in the study (PAL Network, n.d.). Literacy Boost assessment tools have been used in the Doro camp in South Sudan, and EGRA was used in the Kakuma camp in Kenya. These assessments were conducted for specific advocacy or program evaluation purposes, not as part of a national learning evaluation.

Even when national studies are conducted in countries affected by conflict and crisis, children in the most acute emergency situations are likely to be underrepresented. Of the 15 countries affected by conflict or crisis that have reading data on the USAID Early Grade Reading Barometer website,¹³ only four mention the emergency context in the EGRA reports: the 2012 Jordan and 2015 Liberia reports briefly mention the emergency situations affecting these countries' education systems; the 2015 Mali report describes that country's conflict and its potential ramifications for the study results; and the 2015 Philippines report notes that ongoing conflict in the Autonomous Region of Muslim Mindanao required excluding schools in the conflict zone from the assessment. None of the reports disaggregated data by refugee or IDP status or noted efforts to include children affected by conflict and crisis in the study, and none stated that the

¹³ Countries with data on earlygradereadingbarometer.org considered affected by conflict or crisis for this report include Bangladesh, Cambodia, DRC, Ghana, Iraq, Jordan, Kenya, Liberia, Malawi, Mali, Nigeria, Palestine, Papua New Guinea, The Philippines, and Uganda.

samples included EiE contexts or displaced student populations. Another way of addressing security issues was demonstrated with EGRA in Afghanistan, where data collection was suspended in conflicted areas but continued when the security situation improved.

Enabling students to sit for their end-of-cycle examinations is one way UNICEF helps students affected by conflict or crisis continue their education. Negotiating with governments and providing transportation for students are several strategies UNICEF has used in Iraq, Lebanon, Libya, Sudan, Syria, and Yemen (UNICEF, 2018).

Exclusion of EiE contexts is evident in regional and international assessments as well. The 2005 SEACMEQ report for Uganda notes that “schools in areas affected by serious military conflicts and special schools were excluded” (Byamugisha & Ssenabulya, 2005, pp. 33-34) from the study. In Mali in 2012-2013, data collection for PASEC was discontinued in conflict-affected areas, and these geographies were excluded from the study (PASEC, 2014). While not a focus of this study, it is likely that refugee students in high-income countries are also excluded from learning assessments. A 2016 PIRLS report noted that “Austria’s increased exclusions in 2016 resulted from more non-native language speakers, probably due to the refugee crisis in Europe” (Martin, Mullis, & Hooper, 2017, p. 5.27).

Target populations

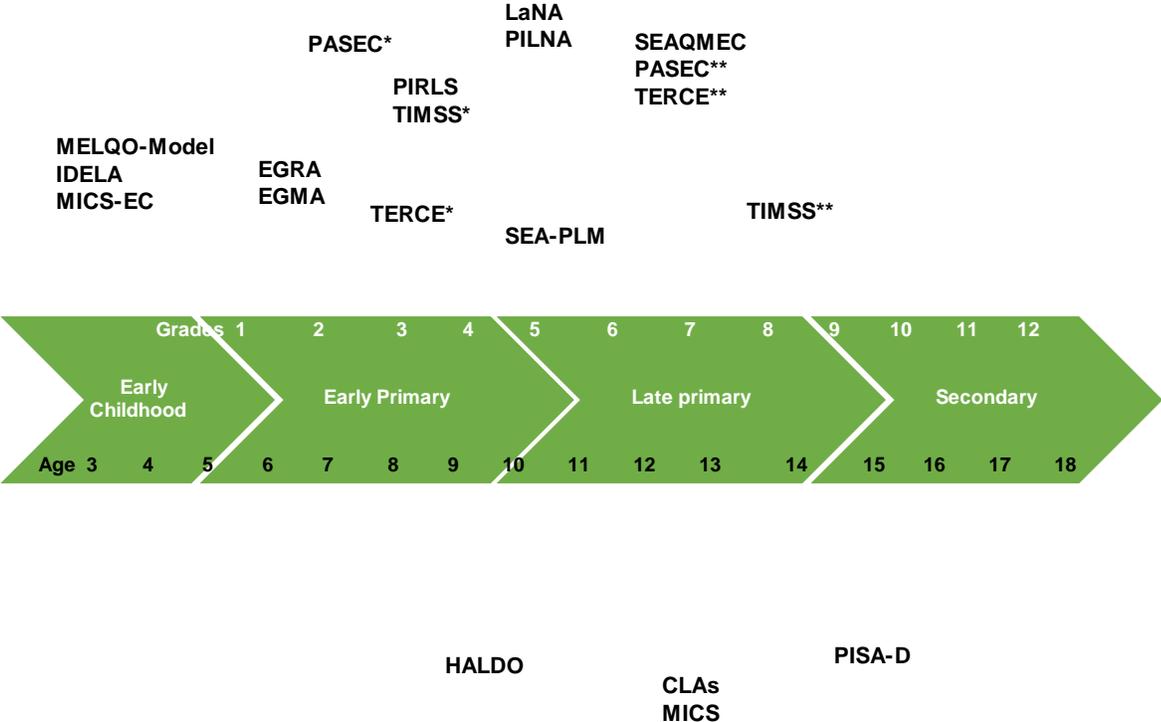
The assessment tools selected for coding target a range of populations, from early childhood to age 18. Annex C contains descriptions of the 30 multi-country assessments included in this exercise. MELQO, IDELA, and MICS-ECDI target early childhood. EGRA and EGMA target early primary grades, while several tools like PASEC, TERCE, and TIMSS have one assessment conducted at an early grade and another at the end of primary school. SEACMEQ, LaNA, PILNA, and SEA-PLM target the last grades of primary school. Several tools that administer assessments outside of school (including MICS, HALDO, and the CLAs) administer the same assessment to children ages 6 to 18. Figure 5 shows the age and grade ranges of the tools.

There are three measurement points for SDG4.1.1: students in grades 2 or 3, at end of primary school, and at end of secondary school. PASEC and LLECE target students in grades 2 and 3, respectively. LANA, PASEC, PILNA, SEACMEQ, and LLECE collect data in grade 6 or at the end of primary school. TIMSS and PISA-D are options for measuring students at the secondary level. One challenge with TIMSS is that significant proportions of the student populations in developing countries are below level 1, meaning that they are not able to correctly answer even the least difficult items on the test. PISA-D shows promising opportunities to assess children and youth with low performance levels.

We mapped cross-national academic learning assessment tools used in 61 countries where EiE is provided. Of the 61 countries, only two—Eritrea and Libya—did not participate in any of the cross-national tools.

The MICS surveys, TIMSS, PIRLS,¹⁴ EGRA, EGMA, MELQO, and IDELA. are international tools, and others cover specific regions. PASEC, SEACMEQ, TERCE, PILNA, and SEA-PLM target Francophone Africa, Southern and Eastern Africa, Latin America, the Pacific Islands, and Southeast Asia, respectively. Several more recently developed tools, PISA-D, HALDO, and LaNA, have been piloted in a few countries, with the intention to expand coverage.

Figure 5. Age and grade ranges of tools included in study



Note: For tools that have two versions of the test, * indicates the early test and ** indicates the later test.

CLAs convened by the People’s Action for Learning Network bring volunteer-administered assessments to 14 countries. CLAs emerged in India in 2005 to raise awareness and advocacy around low learning levels and to promote bottom-up accountability. Volunteers traveled to rural districts to administer simple reading and math tests to any child of primary or secondary school age, whether or not they were in school. The results, which were published in an Annual Status of Education Report (ASER), helped stimulate debate and prioritize learning in national policy. CLAs have assessed 7.5 million children to date and have expanded quickly across the globe, adapting to contexts in Kenya, Tanzania, Uganda, Mali, Senegal, Pakistan, and more. Because CLAs are based on representative samples, they can provide information on hundreds of millions of children.

¹⁴ PIRLS is primarily utilized by high income countries, and only two middle-income countries affected by conflict in our list of countries of interest have participated in the last few rounds: Colombia (2011) and Turkey (2001)

Table 10. Targeted grade or age range of tools in countries where at least one tool has been administered in the last five years

	System-Monitoring Tools				Multi-Purpose Tools		
	Early childhood	Grades 2 or 3	Late primary	Secondary	Early childhood	Grades 2 or 3	Age range*
Afghanistan					⊙	⊙	
Angola			⊙			⊙	
Bangladesh	⊙				⊙	⊙	⊙
Benin	⊙	⊙				⊙	⊙
Burkina Faso		⊙					
Burundi						⊙	
Cambodia			⊙	⊙		⊙	
Cameroon	⊙	⊙			⊙		
Chad	⊙	⊙				⊙	⊙
CAR	⊙	⊙				⊙	⊙
Colombia		⊙	⊙	⊙	⊙		
Congo, DRC	⊙	⊙				⊙	⊙
Côte d'Ivoire	⊙	⊙					
El Salvador	⊙	⊙	⊙	⊙			⊙
Eritrea							
Ethiopia					⊙	⊙	
Gambia	⊙						⊙
Ghana	⊙		⊙		⊙	⊙	⊙
Guatemala		⊙	⊙	⊙	⊙	⊙	
Guinea	⊙	⊙					
Guinea-Bissau	⊙						⊙
Haiti					⊙	⊙	
Honduras	⊙	⊙	⊙	⊙	⊙	⊙	⊙
Indonesia	⊙		⊙	⊙	⊙	⊙	
Iraq	⊙					⊙	⊙
Jordan			⊙	⊙	⊙	⊙	
Kenya	⊙		⊙		⊙	⊙	⊙
Kyrgyzstan	⊙				⊙	⊙	⊙
Lebanon	⊙		⊙		⊙		⊙
Lesotho	⊙		⊙		⊙		⊙
Liberia					⊙	⊙	
Libya							
Malawi	⊙		⊙		⊙	⊙	⊙
Mali	⊙	⊙			⊙	⊙	⊙
Mauritania	⊙						
Mozambique			⊙		⊙	⊙	⊙
Myanmar			⊙		⊙		
Nicaragua		⊙	⊙		⊙	⊙	
Niger		⊙			⊙	⊙	
Nigeria	⊙					⊙	⊙

	System-Monitoring Tools				Multi-Purpose Tools		
	Early childhood	Grades 2 or 3	Late primary	Secondary	Early childhood	Grades 2 or 3	Age range*
Pakistan	⊙			⊙	⊙	⊙	⊙
Palestine	⊙	⊙	⊙		⊙	⊙	⊙
Papua New Guinea			⊙		⊙	⊙	
Philippines			⊙	⊙	⊙	⊙	
Senegal	⊙	⊙		⊙	⊙	⊙	⊙
Sierra Leone	⊙				⊙	⊙	⊙
Solomon Islands			⊙		⊙		
Somalia	⊙				⊙	⊙	
South Sudan	⊙					⊙	
Sudan	⊙				⊙		⊙
Syria			⊙			⊙	
Togo	⊙	⊙					⊙
Tonga	⊙		⊙			⊙	⊙
Tunisia	⊙		⊙				⊙
Turkey			⊙	⊙			
Uganda			⊙		⊙	⊙	⊙
Ukraine	⊙		⊙				
Venezuela		⊙	⊙				
Yemen						⊙	
Zambia			⊙	⊙	⊙	⊙	
Zimbabwe	⊙		⊙		⊙	⊙	⊙

*This generally includes students of primary and secondary school age, but it is not defined by grade because tools are administered outside of the classroom. CLAs, MICS Foundational Skills Module, and HALDO are administered this way, such as using the same tool for ages 4 to 16.

Administration of assessment

How an assessment is administered varies according to context, purpose, and target population. Tools designed for the early primary grades, CLAs, assessments embedded in MICS household surveys, and tools designed specifically for displaced or out-of-school children are usually orally delivered face-to-face by a trained assessor from the administering agency. These occur in a home, school, or community center, depending on the tool. CLAs are primarily administered by trained citizen volunteers in the student’s home.

On the other hand, assessments administered to older children in school (e.g., PISA-D, PIRLS, PASEC) are typically paper based, with multiple-choice and open-response questions. With this type of assessment, results are more likely to be comparable across countries.

The MICS-ECDI is administered to parents as a questionnaire, and the MELQO-MODEL includes parent and teacher questionnaires on the children’s development and learning; there is also a direct assessment option.

Accessibility and costs

The MICS-based tools, CLAs, MELQO, EGRA, EGMA, and IDELA, are open source and available online, but there are costs associated with any in-country implementation. The cost for CLAs is US\$200,000 per testing year on average, depending on how it is implemented (i.e., national or sub-national). All regional assessments and a few international assessments require countries to pay a participation fee in addition to the in-country cost of implementing the assessment. The total cost for regional assessments (e.g., TERCE, SEACMEQ, PASEC), including access to the tool and implementation, ranges from US\$200,000-US\$500,000 per country, depending on assessment program and local costs. The total cost to implement PISA-D, PIRLS, and TIMSS is about US\$800,000 per cycle; variations depend on the country size and context (UIS, 2018a).

The participation fees are US\$225,000 for TIMSS and PIRLS. PISA program fees are around the same, but they vary by country and relate to how much a country pays in OECD membership fees. Fees for regional assessment also vary by country, based on the country's economic status and ability to pay.

Test administration represents about half the total cost for these international assessments, which primarily includes field testing and supervision; institutional costs, which are driven by personnel costs and fees, account for around 25% of the total. Other costs include test preparation, processing and analysis, and dissemination of findings (UIS, 2016). The overall cost of administering EGRA is much lower and is primarily determined by the cost of field testing and personnel (Wagner, 2011).

Assessment construction and purpose

In terms of assessment construction, the majority (80%) of tools included in the mapping are competency based, meaning that they test specific skills. Six of the tools are content based, meaning they are constructed with specific curriculum content in mind. This applies in particular to regional assessments like SEACMEQ, TERCE, and SEA-PLM, and a few others, including TIMSS. Twelve tools are designed for system monitoring; that is, they yield results that are comparable year-to-year and cross-nationally. These include the MICS-ECDI and MICS-FLS household surveys and international and regional assessments. The remaining tools yield results that are context specific and multi-purpose and are not comparable internationally (see Table 11). These include some of the early childhood and early grade assessments such as EGRA and EGMA, HALDO, and the CLAs.

Contextual information collected

Most of the tools gather various contextual and background information along with the assessment results. All tools collect at least information on characteristics like gender, socioeconomic status, language, and education level. Having different sets of information about the characteristics in a particular setting helps guide programming and decisions by providing a more complete understanding of that group's learning challenges or successes. Some tools collect information on student learning outside of school (e.g., time spent reading or studying at home or doing extracurricular activities), household characteristics (e.g., parental education level, mother tongue, socioeconomic status, use of ICT at home),

school characteristics (e.g., school infrastructure and resources, management/leadership, teacher absenteeism, homework), and village/community facilities (health centers, electricity, etc.).

Tools such as PILNA, EGRA, PIRLS, TIMSS, and TERCE often administer background questionnaires to parents, teachers, and principals alongside the learning assessments for children. This provides a more robust picture of the education landscape in each context. In Syria, for instance, the War Stressor Survey, which measures exposure to conflict and the emotional repercussions of trauma, and the Snapshot of School Management and Effectiveness were administered alongside the EGRA and EGMA in order to provide a full picture of primary education in opposition-held areas in Syria. The simultaneous information on academic performance, student well-being, and insights from school staff members allows for analytical insights about the situation in Syria and can support evidence-based design of donor programming (see Box 1). Furthermore, MICS household surveys collect a host of background information on household and family characteristics, including location, health, wealth, disabilities, parental involvement and child labor, which enables a robust analysis of the MICS-ECDI and MICS-FLS data. However, information about refugee or IDP status has not been collected with these tools. Several efforts have been made only recently to gain consensus on how to collect statistics on refugees and IDPs, including by the UIS (UIS, 2018a), UNHCR (2018), and the Education Equity Research Initiative (Soares, Smiley, & Lavan, 2018).

Table 11. Summary of measurement tool characteristics

Assessment	Purpose	Construction	Availability	Method	Site	Context
PASEC	System monitoring	Competency based	Country fee for participation and implementation	Grade 2 oral; grade 6 written multiple choice	School	Background questionnaire includes student, household, and school characteristics
PILNA	System monitoring	Competency based	Country fee for participation and implementation	Written multiple choice	School	Background questionnaire for student, teacher, and principal, including information on school resources and management
PIRLS	System monitoring	Competency based	Country fee for participation and implementation	Written multiple choice and open response	School	Background questionnaires cover student, household, and school characteristics
PISA-D	System monitoring	Competency based	Country fee for participation and implementation	School tool is written multiple choice and open response; out-of-school	School, home (out-of-school component)	Background questionnaires cover student, home, school, and community characteristics

Assessment	Purpose	Construction	Availability	Method	Site	Context
				tool is tablet based		
MICS- ECDI	System monitoring	Competency-based	Open-source	Oral	Home	Comprehensive background information collected as part of general MICS survey
MICS- Foundational Learning Skills	System monitoring	Competency-based	Open-source;	Oral	Home	Comprehensive background information collected as part of general MICS survey. Information collected in the module includes reading habits and home language.
LaNA	System monitoring	Content-based	Not open source	Written multiple choice	School	Background information on school and household characteristics
UNRWA MLA	System-monitoring	Content-based	Not open source	Written multiple choice and open response	School	Background questionnaires on classroom and school characteristics
SEA-PLM	System monitoring	Content-based	Country fee for participation and implementation	Written multiple choice and open response	School	Proposal for background information includes student, household, school, and community characteristics
SEACMEQ	System monitoring	Content-based	Country fee for participation and implementation	Written multiple choice and open response	School	Background information on student, household, and school characteristics; access to books and tutoring
TERCE	System monitoring	Content-based	Country fee for participation and implementation	Written multiple choice and open response	School	Background questionnaires for teachers, principals, students, and families, including characteristics like school management
TIMSS	System monitoring	Content-based	Country fee for participation and implementation	Written multiple choice and open response; online tablet-based option in 2019	School	Background information on school, household, and student characteristics; attitudes toward learning
OLA	Multi-purpose	Competency-based	Not open source	OLA is oral; eOLA uses digital tablets	Nonformal education programs	Demographic information, information on reading, nonformal and formal education background, literacy goals

Assessment	Purpose	Construction	Availability	Method	Site	Context
Literacy Boost	Multi-purpose	Competency-based	Not open source	Oral	School	Background information on household and school characteristics; reading habits at home
Numeracy Boost	Multi-purpose	Competency-based	Not open source	Oral	School	Information about students' exposure to math outside of school
MELQO	Multi-purpose	Competency based	Open source	Oral	School, home, or informal learning center	No standardized collection of background information
STAR	Multi-purpose	Competency based	Open source	Structured task	In or near school	Recommended background includes demographic and economic information, home literacy environment
CLAs*	Multi-purpose	Competency based	Open source; fee for implementation	Oral	Home	Comprehensive background information on demographics, and school, community, child, and household characteristics
EGMA	Multi-purpose	Competency based	Open source; fee or own cost for implementation	Oral	School	No standardized collection of background information
HALDO	Multi-purpose	Competency based	Open source	Oral	School, home, or informal learning center	Background information on demographics, home learning environment, language
IDELA	Multi-purpose	Competency based	Open-source but must sign MOU	Oral	Varies	Information collected on parent education, enrollment in ECE, household assets
EGRA	Multi-purpose	Competency based	Open source; fee or own cost for implementation	Oral	School	No standardized collection of background information

*CLAs include ASER, UWEZO, LearnNigeria, IID/BRAC Survey, SCALE Ghana, TPC Mozambique, Beekunko, and Jangandoo, among others.

Program approaches

Examining particular interventions provides a deeper understanding of the vertical linkages among guidance documents, monitoring and assessment tools, and program design. We reviewed programs from four organizations that focus on strengthening learning and equity in conflict- and crisis-affected locations: Save the Children's Literacy Boost and Numeracy Boost, World Vision's Unlock Literacy, EDC's Read Right Now! and UNRWA's EiE program, which includes the MLA tool. Annex D provides a

description of the program approaches and associated measurement tools.

This snapshot of program approaches demonstrates the tendency of organizations to tweak existing tools to create bespoke assessments that cater to the unique needs of each program. The Literacy Boost assessment tool is similar in many ways to the EGRA tool and Numeracy Boost to EGMA. However, Literacy Boost tests a wider range of emergent reading skills, focuses less on fluency, and collects more detailed student background data, making it better suited to inform evidence-based programming and determine the effects at the lower end of the skill distribution. World Vision's STAR tool similarly draws from EGRA and the Literacy Boost toolkit, but it asks students to identify fewer words and excludes the listening comprehension component. Furthermore, EDC adapted the paper-based EGRA to an online format that was better suited to the need for rapid information, and it created the OLA specifically to capture information on out-of-school older youth.

In each case, an emphasis on learning and equity has taken root in alignment with the global goals. For instance, a recent redesign of the UNRWA MLA assessment was developed in response to an independent review that found that UNRWA education systems emphasized rote learning and memorization, rather than skills acquisition and higher order cognitive abilities. It also found that the UNRWA was constrained by an emphasis on traditional high-stakes examinations that focused on factual content and memorization, which reflected the examination cultures of refugee host countries and the MENA region. The tool has since adopted a holistic approach to assessment that looks at classroom practices, school environments, and equity in student learning outcomes.

Similarly, measurement tools used by Save the Children—the Literacy Boost and Numeracy Boost assessments and IDELA—measure the skills of all students in a program using a continuous scoring system, which supplies details that allow for equity analysis. Paired with surveys that gather information on student background and school context, this enables programs to identify and respond to the unique needs of disadvantaged groups and helps close gaps in learning achievement.

Results of the above programs have generally shown evidence of improvement in terms of academic outcomes. Children in Literacy Boost schools in Malawi and Zimbabwe learned twice as many words as children in similar non-intervention schools, and 34% of Literacy Boost girls in Pakistan could read with comprehension, compared to 8% of girls not in the program (Save the Children, 2013). UNRWA's 2017 Annual Operational Report highlighted an increase in the proportion of students achieving at or above their grade level for almost all content domains and cognitive levels, in addition to notably greater equity in students' learning outcomes (UNRWA, 2018).

To support programs that provide better, clearer evidence of what works to produce the best academic and SEL outcomes for children in fragile contexts, the IRC and TIES/NYU launched the Education in Emergencies: Evidence for Action (3EA) research-practice partnership in 2016. As part of this work, they convened a consortium of research-practice-policy partnerships to develop, adapt, and test a set of measures to assess critical dimensions of program implementation quality and children's learning and holistic development in crisis contexts. These measures comprise the INEE MENAT Measurement Library on the INEE website. The library will also offer training and guidance materials to facilitate its use,

including a decision-making tree that helps users understand the questions they should be asking themselves as they plan their assessments and an inventory of measurement tools with evidence of use in the region.

Box 1. Example EiE case: Administration of ASER, EGMA, and EGRA in Syria

There is little concrete evidence about the learning levels of children in Syria, but some effort has been made to gather such information in the last few years. As donors and humanitarian assistance organizations strive to understand the situation in Syria and design education programs that meet the needs, assessment tools like ASER, EGRA, and EGMA are being used to collect information on learning.

The IRC piloted the ASER tool, a citizen-led assessment launched in 2005 to document learning levels in India, in the Syrian context. ASER was chosen because its tools are easy to administer and they provide a quick snapshot of students' reading and math abilities. The initial pilot took place in December 2015, when the test was administered to 122 students in grades 1-8 in two schools. The ASER assessment was subsequently administered in November 2016 to 2,5846 children in five IRC-supported schools in Idleb. The IRC education team trained the teachers who conducted the assessment using paper forms over a two-day period. The IRC team then entered the data electronically. In general, scores improved for students in higher grades, but older students still had low scores. More than half of 6th and 7th graders could not read a simple, 7-10-sentence story—the equivalent of 2nd-grade reading skills. About 64% of 6th graders, 63% of 7th graders, and 46% of 8th graders could not solve a subtraction problem—the equivalent of 2nd-grade math skills. Security challenges and children's obligations at home made it impossible to reach all of them; the final number assessed represents 73% of the children enrolled in grades 1-8. **This**

process provided key insights not only into learning in Syria but also into the feasibility of measuring learning in difficult contexts.

In addition to ASER, the UK Department for International Development supported the administration of EGRA and EGMA tools to 1,500 grade-3 pupils in the Syrian context (Rural Damascus, Aleppo, Idlib) in April-May 2017. Of the grade-3 pupils assessed, most had not yet developed the level of reading or math skills expected at that grade level. They also had particularly weak decoding skills and difficulty with more complex calculations. The War Stressor Survey (understanding how children have internalized conflict) and Snapshot of School Management and Effectiveness (SSME) surveys were administered simultaneously in order to build a full picture of early primary education in opposition-held areas in Syria. The SSME tool revealed that more than half of the teachers surveyed had no training in pedagogy and 90% of the schools lacked reading materials. The War Stressor Survey revealed the effects of exposure to conflict and the emotional repercussions of trauma. It showed that more than 60% of the children surveyed had experienced shelling or bombing and more than half had lost a family member in the war. As a result, more than 80% of the children suffered from constant anxiety. **The combination of data on academic performance, student well-being, and insights from school staff members provided analytical insights about the situation in Syria and supports evidence-based design of donor programming.**

Save the Children Syria contributed information for this description. Other sources used:

International Rescue Committee. 2017. *Report: Impact of war on Syrian children's learning: testing shows gaps in literacy and math skills*. Retrieved from www.rescue.org/report/impact-war-syrian-children-learning-testing-shows-gaps-literacy-and-math-skills

UKAid & Chemonics. 2019. *Improving early learning outcomes in opposition-led Syria*. Retrieved from www.heart-resources.org/wp-content/uploads/2018/04/DFID_improving-early_learning-outcomes_syria-1.pdf

Part 3. Domain Mapping: Literacy and Numeracy

While all assessments included in the analysis measure some understanding of basic math and/or literacy, each tool varies in the particular skills, knowledge, and competencies children are required to demonstrate. While concepts of literacy and numeracy may appear conventional on the surface, measures of learning are highly varied and tools differ in what determines achievement. The next section illustrates commonalities and differences across the domains and sub-domains assessed by each tool.

Coding system

To enable comparisons across the different tools, we adopt the framework and coding scheme created by the UNESCO IBE and UIS to map content domains and sub-categories in mathematics (UNESCO IBE & UIS, 2017) and reading (UNESCO IBE & UIS, 2018). The coding scheme provides a clear picture of the measurement of academic content and competencies, with the aim of linking different assessments and categorizing standards, contents, skills, and performance levels in a globally comparable way. With this methodology, we are able to identify commonalities and disparities across content area and coverage in the measurement of academic skills.

The coding scheme was derived from studying national assessment frameworks (NAFs) in English, French, and Spanish that ranged from primary to lower secondary education to align with the three points of measurement of SDG indicator 4.1.1: grades 2 and 3, end of primary school, and end of lower secondary school. To best understand what reading content countries are assessing, UNESCO IBE/UIS examined 73 NAFs from 25 countries; for math content, 115 NAFs from 53 countries were assessed and catalogued.¹⁵ The NAFs collected were from a mix of high-, middle-, and low-income countries representing most world regions.¹⁶

The coding scheme for mathematical content is comprised of 6 domains, which are then broken down into 17 sub-domains; the reading content coding scheme has 3 domains and 6 sub-domains. As the EPWG is interested in literacy (reading and writing) and not just reading outcomes, an additional domain of “writing” was added to the UNESCO IBE/UIS coding scheme. The UNESCO IBE/UIS mathematics coding scheme already includes elements of numeracy, so no additional domains were added. Figure 6 and Figure

¹⁵ Many countries conduct national assessments at more than one grade level, which results in a higher number of NAFs.

¹⁶ The only region not represented in the framework is Central Asia, due to language limitations. For reading component, NAFs were analyzed from the following countries: Egypt (1), Qatar (2), Bosnia and Herzegovina (1), Estonia (2), Canada (7), England (2), France (3), Ireland (2), Spain (2), Gambia (4), Mauritius (1), Senegal (2), Seychelles (8), Honduras (9), Mexico (7), Peru (3), Australia (3), Cambodia (2), Hong Kong SAR (2), Micronesia (2), New Zealand (6), India (2), and Pakistan (2). For math, Belgium (3), Canada (9), France (3), Iceland (2), Ireland (2), Malta (1), Spain (2), England (2), USA (2), Australia (3), Cambodia (2), Fiji (3), Hong Kong SAR (2), Laos (1), Micronesia (3), New Zealand (8), Philippines (1), Bangladesh (1), India (3), Nepal (1), Sri Lanka (2), Argentina (2), Chile (2), Colombia (1), Costa Rica (1), Dominica (6), Dominican Republic (1), Ecuador (1), El Salvador (1), Guatemala (3), Guyana (2), Honduras (8), Mexico (6), Peru (2), Saint Vincent and the Grenadines (2), Jordan (1), Egypt (1), Pakistan (2), Palestine (1), Albania (2), Estonia (2), Gambia (2), Ghana (2), Ivory Coast (1), Mauritius (1), Senegal (1), Seychelles (3), South Africa (5), South Sudan (1), Uganda (2), and Zambia (1)

7 illustrate each domain and sub-domain for numeracy and literacy, respectively, and descriptions of their component parts are provided in Annex E.

Other academic domains such as science, civics and citizenship, and arts and culture are referenced in the analysis, but the quantitative mapping process focuses solely on literacy and mathematics.

Mapping literacy domains and sub-domains

The literacy-specific EGRA measurement tool and the EDC OLA are the most comprehensive tools reviewed, meaning that they assess the widest breadth of literacy sub-domains.

The entire set of EGRA sub-tasks covers all domains in the UNESCO IIEP/UIS coding system at the most basic level, although countries rarely use all sub-tasks. The EGRA toolkit provides guidance on how to bring together local experts and stakeholders to adapt and develop the tool in any context. Core EGRA sub-tasks include letter naming and sound identification (phonological awareness), non-word reading (decoding), listening comprehension (listening), familiar words (fluency), and paragraph reading (fluency and comprehension). Additional sub-tasks used less commonly include dictation (writing) and phoneme segmentation (pronunciation/speaking).

OLA also provides an example of a short (~20 minute), simple, yet comprehensive literacy measurement tool. It was developed to measure the literacy skills of youth and adults in post-conflict environments and extreme poverty who have minimal literacy acquisition, and is currently used as an evaluative and diagnostic tool for youth enrolled in non-formal education programs. OLA orally assesses foundational reading skills like letter sounds (phonological awareness), letter and syllable identification (decoding), text reading and comprehension (comprehension), and real-life reading (vocabulary). An assessment of basic writing (in the pilot stage) and an oral language comprehension assessment (listening) are planned.

For paper-and-pencil tests such as the regional and international tests, the linguistic competency domain is rarely assessed, and in some cases more basic competencies like decoding and phonological awareness are also excluded. Reading comprehension, including a range of literary styles and levels of technicality, is typically the main focus. The PILNA, TERCE, and SEA-PLM regional assessments cover a large majority of sub-domains but exclude listening, speaking, and direct testing of decoding. PIRLS, the international comparative assessment measuring student learning in reading in the 4th grade, also does not cover linguistic and metalinguistic competencies, but it does assess reading comprehension and writing in great depth. PISA-D also focuses mainly on reading competency and includes few other literacy domains and sub-domains.

Figure 6. Literacy domains and sub-domains

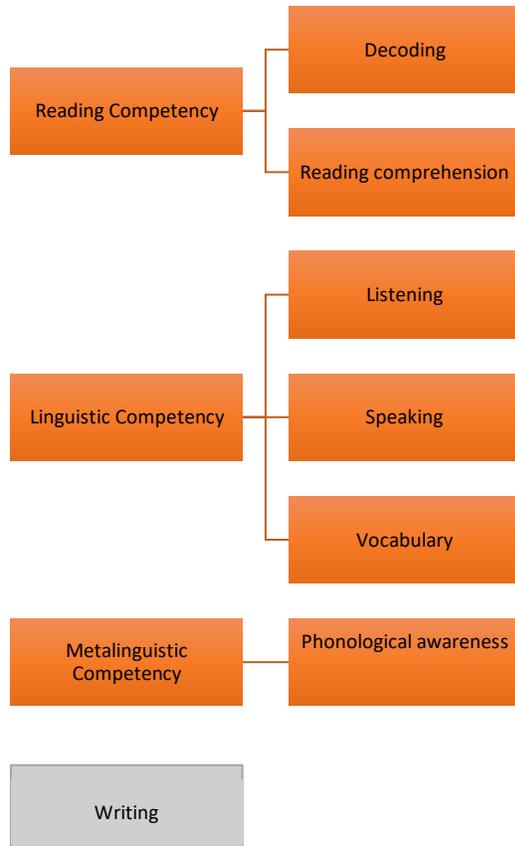
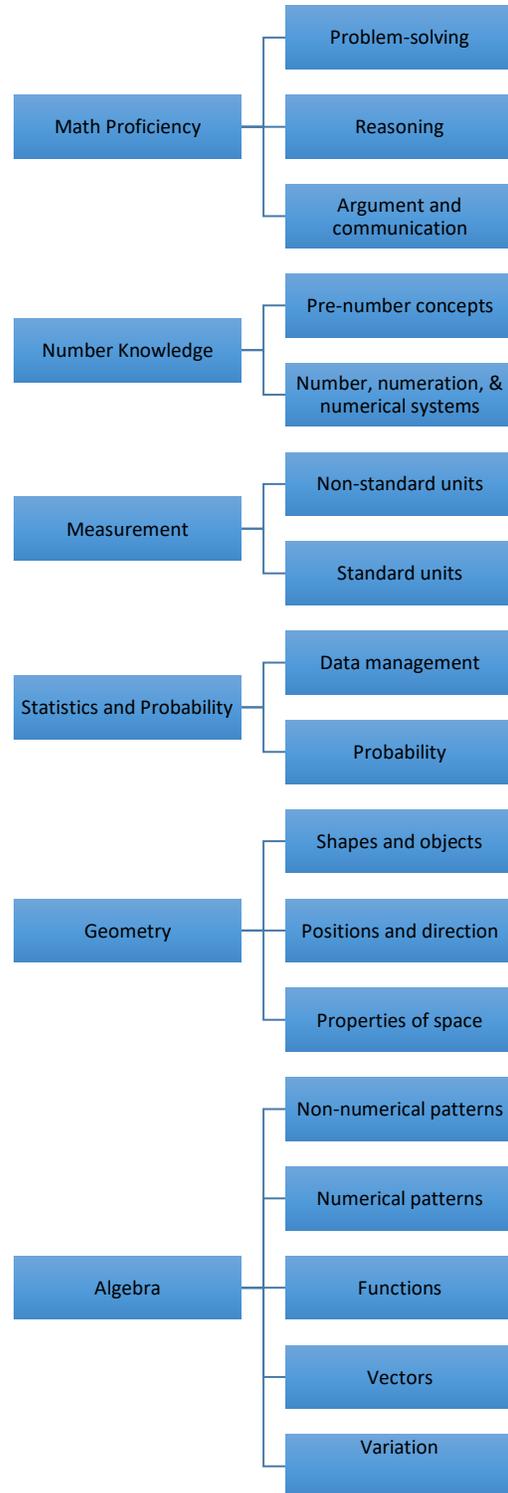


Figure 7. Numeracy domains and sub-domains



The eight CLAs reviewed evaluate a narrower set of literacy components, all within the reading competency domain. All CLA tools cover decoding, and most contain a couple of questions on reading comprehension on a simple passage. HALDO, the EiE-specific tool, covers decoding and comprehension and also includes vocabulary—for example, having the child say the names of all the animals they know.

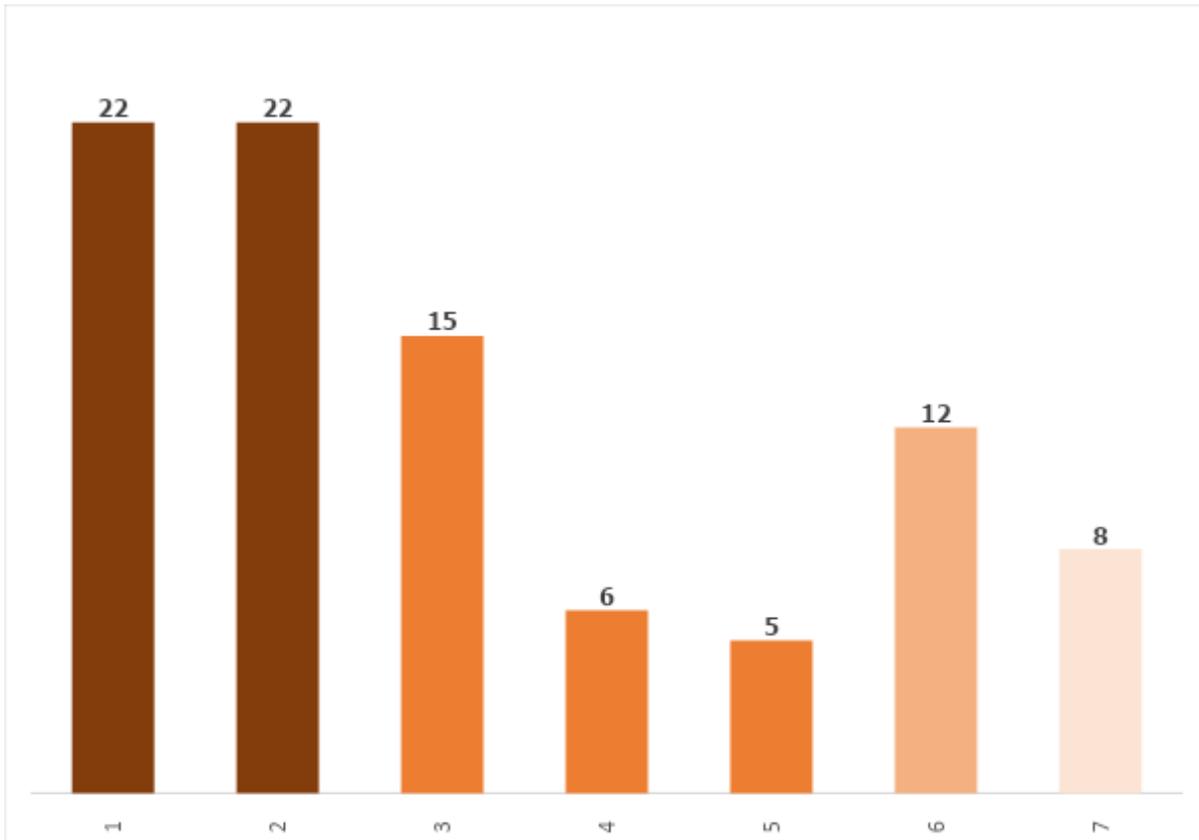
Table 12. Types of assessment tools by literacy domain and sub-domain

Category	Tools	Reading		Linguistic			Metalinguistic	Writing
		Decoding	Comprehension	Listening	Speaking	Vocabulary	Phonological awareness	
System-monitoring tools								
International assessments	PISA-D, TIMSS, PIRLS, LaNA		Darkest blue			Lightest blue	Lightest blue	Lightest blue
Regional assessments	LLECE, SEACMEQ, PASEC, PILNA, SEA-PLM, UNRWA MLA	Lightest blue	Darkest blue	Lightest blue	Lightest blue	Darkest blue	Darkest blue	Lightest blue
Household	MICS-ECDI, MICS-FLS	Darkest blue	Lightest blue			Lightest blue	Lightest blue	
Multi-purpose tools								
Foundational skills	MELQO, EGRA, STAR, Literacy Boost	Darkest blue	Lightest blue	Lightest blue	Lightest blue	Lightest blue	Lightest blue	Lightest blue
Citizen led	ASER, UWEZO, LearnNigeria, etc.	Darkest blue	Darkest blue					
Designed for EiE contexts	HALDO, IDELA, OLA	Darkest blue	Lightest blue	Lightest blue	Lightest blue	Darkest blue	Lightest blue	Lightest blue

-  Darkest blue indicates that all tools in the category include sub-domain
-  Lighter blue indicates that half or more tools in the category include sub-domain
-  Lightest blue indicates that one or a few tools in the category include sub-domain
-  No color indicates that no tools in the category include sub-domain

The ability to read and comprehend a simple text is understood as a fundamental skill across the globe and is the foundation children need to succeed in other areas of education. As children pass through the grade levels, more and more academic content is transmitted to them through text, therefore, their ability to learn depends on their ability to read, which is also essential to a person’s self-guided learning outside the classroom. Therefore, it is logical that reading competency (decoding and comprehension) is by far the most prevalent domain in the review.

Figure 8. Literacy sub-domains by number of tools



Note: Total N= 27 tools (excluding 3 numeracy-only tools from overall total)

Reading competency—the capacity to decode and understand written texts—is the most prevalent domain of literacy. All assessments reviewed, except the two numeracy-specific tools (TIMSS and EGMA), test at least one sub-domain within the reading competency domain. The two sub-domains of reading competency are decoding and reading comprehension, which are more or less equally prevalent in the set of tools.

Decoding is a foundational reading skill, defined in the UNESCO IBE/UIS coding scheme as the ability to associate the orthographic form of a word with its phonological form, where the orthographic form is given by the sequence of graphemes. This includes constructs like the alphabetic principle, precision in reading words and sentences, pronunciation, and fluency. Because decoding is a fundamental skill and can be assessed most easily orally, it was found to be included in tools targeting earlier grades, and in CLAs, household survey-based assessments, and tools that specifically target displaced populations, all of which are administered orally.

Reading comprehension is the process by which information is retrieved from a written text, interpreted, and reflected on. Most tools, other than those targeting early childhood, included some form of reading

comprehension, from questions based on a sentence or short story in HALDO, to a more technical or complex text in PISA-D and the regional assessments targeting later grades. While some CLAs do include comprehension questions, about half simply test the ability to read rather than comprehension. In these assessments, comprehension is sometimes assumed if/when fluency, accuracy, and prosody in the reading of a passage suggest that the child is understanding. As would be expected, beginner tasks such as decoding are more frequently found in assessments of early childhood and in early grades, while more complex tasks such as reading comprehension are found more frequently in assessments at higher grades.

Within the **linguistic competency** domain, vocabulary was the most commonly included sub-domain, with about 15 tools testing some form of vocabulary knowledge. This sub-domain includes knowledge of words and their meaning, including assessing through definition, context, and images; identification of words using synonyms, antonyms, family words, and categories; and distinguishing linguistic varieties and the elements that make up a word. Vocabulary was most commonly tested in tools targeting the youngest population group, such as very basic word identification using images, but also in the regional assessments and EGRA, which test a more comprehensive set of literacy sub-domains. Speaking and listening were directly assessed in a few tools (5 and 6 tools, respectively), such as EGRA and some program approaches, all administered orally, but they were rarely included in the regional and international written tests.

Twelve tools assess **metalinguistic competency**, which is comprised of the phonological awareness sub-domain. This is the ability to identify phonological units, link them to compose a word, and divide words into phonological units. This can include rhyming, assessing knowledge of syllables, spelling, and compound words. The subset of tools that assess phonological awareness are similar to those that assess vocabulary, such as those targeting the early years, regional assessments, and comprehensive literacy tools like EGRA.

Writing was assessed in some of the regional assessments, as well as PIRLS, EGRA, and two early childhood tools, IDELA and MELQO. An additional writing component is being developed for several tools.

Mapping numeracy domains and sub-domains

The list of mathematics sub-domains is quite long. No tools cover all sub-domains, and tools that cover more advanced content, such as vectors and probability, are particularly sparse. PISA-D covers the greatest breadth of mathematics material, including all sub-domains except the most basic concept (pre-number ideas) and one of the most advanced concepts (vectors). Other regional and international tools that have broad coverage in mathematics are TIMSS, TERCE, SEACMEQ, and SEA-PLM. Apart from number knowledge, content in these tools generally includes measurement, basic data tables, problem-solving, and geometry.

Assessments focused on early childhood and early grades target a narrower set of sub-domains and at a more foundational level. IDELA and MELQO-Model track children's problem-solving processes and

communication, as well as pre-number ideas, measurement with non-standard units, and identification of basic shapes. EGMA covers problem-solving and reasoning, numbers, and number patterns.

Unlike the longer classroom-based assessments, rapid assessments like HALDO and the CLAs focus singularly on numeracy. The only sub-domains included in these kinds of tools are numbers and number systems. In HALDO, the numbers task involves asking the child to select 3, 7, and 15 items (like beans or stones) from a pile, basic addition and subtraction, and word problems. Several CLAs incorporate culturally appropriate math problems. For instance, Uwezo Kenya has a section for ethno-math, including prices (in Kenyan shillings) of everyday objects, like mangos and milking jelly.

Table 13. Types of assessment tools by numeracy domain

Category	Tools	Math Proficiency (e.g., problem-solving, reasoning)	Number Knowledge (including operations)	Measurement	Statistics and Probability	Geometry	Algebra
System-monitoring tools							
International assessments	PISA-D, TIMSS, PIRLS, LaNA	Lightest blue	Darkest blue	Lightest blue	Darkest blue	Darkest blue	Lightest blue
Regional assessments	LLECE, SEACMEQ, PASEC, PILNA, SEA-PLM, UNRWA MLA	Darkest blue	Darkest blue	Darkest blue	Lightest blue	Lightest blue	Lightest blue
Household	MICS-ECDI, MICS-FLS	No color	Darkest blue	No color	No color	No color	Lightest blue
Multi-purpose tools							
Foundational skills ¹⁷	MELQO, EGMA, Numeracy Boost	Lightest blue	Darkest blue	Lightest blue	No color	Lightest blue	Lightest blue
Citizen led	ASER, UWEZO, LearnNigeria, etc.	Lightest blue	Darkest blue	No color	No color	No color	No color
Designed for EIE contexts	HALDO, IDELA	Lightest blue	Darkest blue	Lightest blue	No color	Lightest blue	No color

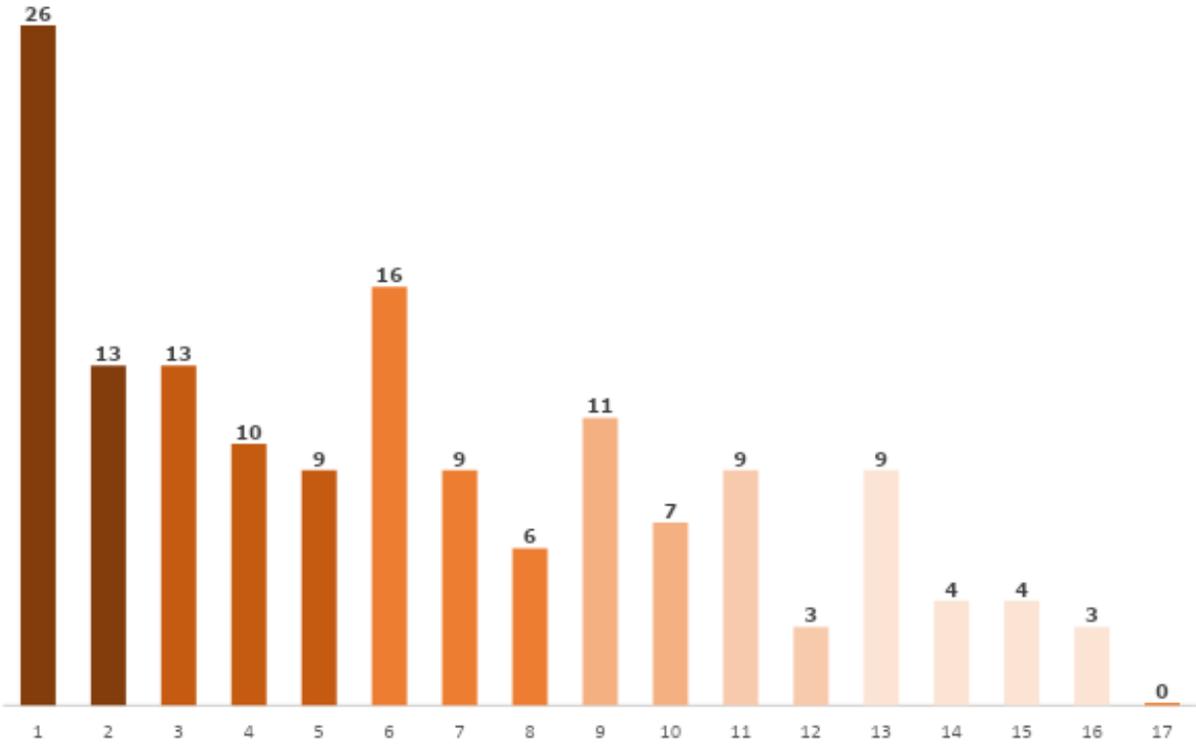
-  Darkest blue indicates that all tools in the category include sub-domain
-  Lighter blue indicates that half or more tools in the category include sub-domain
-  Lightest blue indicates that one or a few tools in the category include sub-domain
-  No color indicates that no tools in the category include sub-domain

All societies use mathematical knowledge to keep track of things, trade, and sell, etc. Citizenship (e.g., knowing basics about the economy, taxes, voting), finances and loans, and personal health maintenance

¹⁷ Tools measuring early childhood and early primary grade competencies.

require basic math skills. Spatial knowledge is used to remember locations and distance. Along with numeracy, spatial knowledge and measurement are some of the most broadly accepted global conventions in math, and they require few language skills. Outside of these basic conventions, much of mathematics does require a special vocabulary, which often is language or dialect specific and is in large part culturally constructed (Gay & Cole, 1967; Nunes & Bryant, 1996; Saxe, 1991; Vygotsky, 1978; Perry, Young-Loveridge, Dockett, & Doig, 2008.)

Figure 9. Numeracy sub-domains by number of tools



Total N=26 tools (excluding 4 literacy-specific tools)

It is logical that **number knowledge** is the most prevalent domain in the mapping. The sub-domain of numbers and number systems is by far the most prevalent. Every tool reviewed included some form of it: counting, basic number knowledge, describing numbers (equal to, less than, greater than), other addition/subtraction, multiplication/division, fractions, decimals, exponents, and alternate number systems (e.g., Mayan numeration). The most rudimentary assessments like those that target early childhood include at least some form of counting, and the most advanced, like TIMSS, include fractions, decimals, and exponents. In the IID/BRAC survey and ASER, the numbers and number systems sub-domain is the only one tested, because these two tasks are simple tests of basic numeracy.

Pre-number ideas involves working with concrete objects to test number knowledge, for example, by counting concrete objects, grouping like objects, or adding and taking away objects to practice simple addition/subtraction sentences. Most of the tests targeting early childhood and the early primary grades include numeracy with concrete objects or images. A few CLAs and the Numeracy Boost program approach also include this sub-domain.

Many tools included precursor skills to **geometry**, such as testing identification of simple shapes and objects found in everyday life. Early childhood tools, two CLAs (in Mali and Senegal), and the Numeracy Boost program approach assess this. TIMSS, PISA-D, TERCE, and PASEC assess more advanced geometry skills, like the position and direction—translating, rotating, and dilating various geometric shapes—and properties of space—locating shapes on the Cartesian plane.

Within the **mathematics proficiency** domain, problem-solving is the most prevalent sub-domain and it is covered in all of the regional and international mathematics tools. Problem-solving is described as the ability to plan, do, and check work during the solution of a mathematical problem. Some of these tools also cover the reasoning sub-domain, and a few, including PISA-D, TERCE, and SEA-PLM, cover argument and communication. The reasoning sub-domain includes personal approaches to problem-solving, constructing diagrams and accounting for initial conditions, estimating before computing, or explaining possible causes of errors. The communication sub-domain involves using mathematical vocabulary to explain procedures and results—orally, visually, and in writing—and communicating information using everyday life examples.

The **measurement** domain is divided into sub-domains of standard and non-standard units. Standard units involves measuring quantities in the world or in daily life using established measurement units (e.g., inches, gallons, pounds, months, days, hours, minutes, currency, temperature). Most regional and international tests at the higher grades assess measurement using standard units. A few CLAs (in Mali and Senegal) and Numeracy Boost also cover standard unit measurement. Non-standard units involve measuring quantities in the world or in daily life using locally derived or estimated units (e.g., book length, stone weights, estimates of time duration). This is primarily covered in some of the early childhood and early grade assessments, but also in the Numeracy Boost tool and in PISA-D.

Nine tools cover the **statistics and probability** domain, mainly with the data management sub-domain, which usually involves reading a simple data table and interpreting the results. These are all regional and international pencil-and-paper assessments. The probability sub-domain is covered in only three tools—PISA-D, TIMSS, and SEA-PLM.

Algebra is the least prevalent domain. The two most prevalent sub-domains in the algebra domain are numerical patterns and sequences, and non-numerical patterns. Non-numerical patterns includes investigating patterns in the environment (e.g., colors, shapes, and sounds) or patterns in cultural activities (e.g., handicrafts, music). Numerical patterns are assessed in several of the regional and international assessments (TIMSS, PISA-D, SEA-PLM, PILNA). IDELA and TERCE tests include non-numerical patterns. Functions (linear and non-linear) and variation (ratio, percentage, and proportions) are included

in a few of the written assessments that have older target populations, such as TIMSS and PISA-D. None of the tools include the vectors sub-domain.

National examinations and classroom assessments

Types of tools that measure learning around the world vary in formality, cost, and time. The majority of assessment tools that were accessed and coded in this mapping are international or regional assessment surveys, CLAs, or assessments associated with particular programs. Ways of measuring learning globally include high-stakes national examinations and more informal classroom assessments, both of which are used to measure learning in EiE settings. However, we were unable to include them systematically in this mapping, given their inaccessibility (sharing national examinations could allow cheating) and the sheer number of classroom-based assessments used worldwide. A short discussion of these assessments and a few examples of their use in EiE settings follows.

National examinations, such as completion/promotion exams, matriculation exams, or A-levels in the UK system, are the most formal, high-cost, and time-consuming standardized tests. Many humanitarian assistance actors are using national assessments as monitoring tools for programming. For instance, the Uganda National Assessment of Progress in Education, a curriculum-based grade-3 and grade-6 literacy and numeracy exam, is used as an indicator of student learning for UNICEF projects. The Norwegian Refugee Council's Accelerated Education Programs in the Democratic Republic of Congo is also using national assessments to check progress, following the guidance of the Accelerated Education Working Group.

National exams test competencies in accordance with a country's national curriculum. Several examples of literacy content in national assessments include a comprehensive range of domains and sub-domains that is similar to regional and international assessments. Listening, reading comprehension, and grammar are included in the Ghanaian national assessment, while writing comprehension and phonological awareness are included in the Cambodian assessment; both of these exams also include a wide range of numeracy domains. In Ghana, the grade-4 and grade-6 national assessment in 2016 included number operations, measurement, geometry, data management, and probability (RTI International, 2016). In Cambodia's grade-6 national assessment, domains included number knowledge, measurement, geometry, algebra, and statistics (Chinna, 2016). It is important to recognize that the purpose of public examinations is to pass or fail students and thereby determine who advances to the next education level. Therefore, while examinations are a major component of education systems in many countries, their use must be carefully evaluated in situations where they might lead to further exclusion or stigmatization of students.

The GAML's technical work on SDG measurement indicates that national assessments can be adjusted and equated using content alignment, a statistical process, and international assessments, but there are still challenges related to reliability and comparability. The GAML content alignment exercise developed a basic mapping of national assessments of literacy (25 countries) and numeracy (53 countries) domains

and sub-domains. The most prevalent sub-domains for reading in that study are similar to the findings in this exercise. All national exams reviewed included reading comprehension tasks. At the primary level, about half of the exams covered decoding, listening, and speaking; a third of the exams at the lower secondary level covered these sub-domains. Vocabulary is hardly covered at all in the national examinations at grades 2 and 3, but it is covered in all the exams at the end of primary and about a third of the exams at the lower secondary level. Number knowledge and geometry are included in all of the mathematics national examinations, most of which also include measurement, statistics, and probability. Testing mathematics proficiency (problem-solving, reasoning, communicating) is much less common, and algebra is included in the examinations targeting high late primary and lower secondary school.

Classroom assessments tend to be low cost, more informal, less time-consuming, and lower stakes. They include formative assessments, which enable teachers to determine whether learning is taking place, such as homework, quizzes, and presentations. They also can be summative, like checking mastery of certain types of information at the end of the unit or term. Although these tests are often ad hoc and not designed for comparability, they are important to consider, as they may be the only form of learning assessments done in many contexts. In Iraq, the Norwegian Refugee Council uses daily 5- to 10-minute activity-based assessments of literacy and numeracy, which are built into scripted lessons. An NGO school for Afghan refugees in Greece uses a variety of assessments of phonetics and math (see Box 2).

Box 2. Classroom assessments for Afghan refugees in Central Greece

In June 2016, the Oinofyta Community School was founded in a refugee camp in Central Greece as a project of the NGO ArmandoAid. Serving only Afghan refugee children, the priority of the school was to settle pupils into a safe and happy learning environment. Students ranged in age from 5 to 15, their education levels differed significantly, and the school population was transient, with new students coming in every day. The population of the school fluctuated from 100 to 400 students during its 18-month operation; one-third of the school population immigrated to Serbia with their families during this period.

The Oinofyta curriculum included literacy, numeracy, and some other subjects. English lessons were based on the “letters and sounds” phonics program, which was adapted for ESL students. Mathematics was delivered using Montessori methods and the Singapore math method. Gardening was used to introduce students to the life cycles, the environment, and healthy eating. Health-awareness campaigns, including habits like teeth-brushing, were included in the curriculum, as were geography and music. When students were absent, new to the camp, or falling behind in any subject area, small-group catch-up sessions

were provided in the afternoons, as formal school hours were in the morning.

Classroom assessments were used to monitor each child’s learning, behavior, and general performance. Each child was assessed for English phonetics using a checklist covering the recognition of 44 English sounds. The school also administered individual monthly reading assessments, including an evaluation of reading strategies used, reading strengths, and next steps for improvement. For math, the students were individually evaluated on different domains, including numeracy and number operations, geometry, measurement, and algebraic equations. The behavior and socio-emotional skills of each child were also evaluated regularly. The information on learning was used to improve teaching and operation of the school, and it was shared with donors and other development partners to highlight progress and challenges.

Maria Siu Munro, founder of ArmandoAid, contributed information for this description in response to an INEE survey in June 2019.

Part 4. Analysis and Conclusions

Application of guidance documents and measurement tools for EiE contexts

Although the tools listed here are used in crisis-affected countries, only three tools—HALDO, OLA, and UNRWA MLA—were designed specifically for emergency contexts and to target displaced populations. HALDO was developed for rapid deployment with minimally trained assessors in the immediate onset of displacement. While the tool measures literacy, numeracy, SEL, and executive functioning, it is not a comprehensive measure of domains, and it uses dynamic scoring to assess children’s skill levels from emergent to advanced. Like the CLAs, it focuses on a wider age range to account for varied skills among children living in emergency contexts.

Some of the tools are simple and flexible enough for administration in conflict-affected areas, even though they may not have been designed specifically for such contexts. These include CLAs, MICS, and several program approaches. Characteristics of these tools that make them viable for EiE contexts include the minimum training required for the assessment administrators and the fact that the administrators can be volunteers and community members, flexible assessment location (household, school, community center), a broader age range, inexpensive administration, short assessment time, and oral-based assessing, for which few physical materials are required. EGRA and EGMA are also flexible in terms of the assessment location, have a short assessment time, are less expensive to administer, and require a shorter training time than international and regional assessments. At the same time, these simpler tools monitor a smaller set of skill domains and constructs. They primarily cover reading comprehension and, to a lesser extent, decoding, except for EGRA, which covers all literacy domains and sub-domains. The math tools that are simpler and more flexible focus primarily on number knowledge (pre-number ideas, counting, basic number operations). Very basic geometry, measurement, and algebra are covered by some program approaches that are used in or designed for EiE settings.

Regional assessments and some international assessments that are not specifically designed for EiE contexts are more formal, cost more, take more time, require more preparation, and involve taking stock of test-takers before they are administered. The advantage is that the results of these assessments can be compared across countries and over time, and they sometimes capture a more comprehensive set of learning domains. For literacy, this tends to include reading comprehension and other sub-domains, such as vocabulary, phonological awareness, and writing. For numeracy, the domains included often go beyond number knowledge to include mathematics proficiency like problem-solving and reasoning, measurement, geometry, and statistics and probability. Algebra is included in some, but to a lesser extent. National examinations are formal and cost more, and they are already in operation in many countries. Because they are based on the national curriculum, they often have a broader domain content as well, but with a core focus on the main domains of reading comprehension and number knowledge.

Connections between local, national, regional, and international tools and frameworks

There are some linkages among international, regional, national, and local frameworks and tools. As the agreed-to global reference point, the SDGs and related indicators are widely influential, and many regional and organization-specific learning frameworks align with SDG guidance. Some regional frameworks define themselves as a regional extension of the SDGs, thereby owning the goals and adapting them to their regional aspirations. Several frameworks, including the Continental Strategy for Education in Africa, use the SDG indicators for measuring learning.

According to the UIS, nine cross-national learning assessments meet criteria to measure SDG indicator 4.1.1, the proportion of young people achieving minimum proficiency in reading and mathematics in grades 2 and 3, at the end of primary and the end of secondary school. These agreed-to assessments are LaNA, PASEC, PILNA, PIRLS, PISA, SEACMEQ, SEA-PLM, TERCE, and TIMSS. Although these are the official assessments listed for measurement, a new methodology to monitor SDG4.1.1 progress that involves agreement on minimum proficiency levels may allow CLAs and national assessments to be included in SDG4 monitoring. However, there are both technical and political challenges in linking and comparing the content of national or local tools in a reliable way. Defining a minimum level of competency as a worldwide standard may face pushback from country officials who emphasize differences in learning and definitions of literacy and numeracy in varied education contexts. There also may be political consequences because of how much of the national curriculum is represented in the definition of minimum proficiency levels (Treviño & Órdenes, 2017).

There is some evidence that governments affected by conflict and crisis are improving their assessment systems with the help of international organizations. GPE recently conducted an internal thematic review of the Partnership's support to learning assessment systems. The review focused on 36 education sector plan implementation grants that were active in 2018, 27 of which meet the definition of conflict- or crisis-affected for the purposes of this paper. The review examined the barriers, needs and challenges related to learning assessment systems noted by countries in their education sector analyses and education sector plans. It also examined the planned activities related to learning assessment systems found in the education sector plans and the implemented activities found in the education sector plan implementation grant documentation.

The review found that more than half (14) of the GPE grant recipients that are conflict- or crisis-affected identified barriers, needs and/or challenges related to learning assessments in their education sector planning and education sector analysis documentation. Nearly all (25) planned to conduct activities related to strengthening their learning assessment systems using GPE grant funds, and 22 ended up using their GPE funding for these types of activities. Countries sometimes shift priorities from the planning to the implementation phase for a variety of political and contextual reasons. Sometimes planned activities end up being funded by other sources.

Table 14 Conflict- or crisis-affected GPE grant recipients prioritizing learning assessment

Conflict- or crisis-affected GPE grant recipients (out of a total of 27) that:		
Identified barriers, needs and/or challenges related to learning assessments (14)	Planned for activities related to learning assessment (25)	Used GPE funding to support some type of learning assessment (22)
Cambodia	Burkina Faso	Burkina Faso
Democratic Republic of Congo	Burundi	Cambodia
Ethiopia	Cambodia	Cameroon
The Gambia	Cameroon	Chad
Guinea	Chad	Cote d'Ivoire
Liberia	Democratic Republic of Congo	Democratic Republic of Congo
Malawi	Eritrea	Eritrea
Mozambique	Ethiopia	Ethiopia
Nigeria	The Gambia	The Gambia
Pakistan (Balochistan)	Guinea	Guinea
Somalia (Puntland)	Kenya	Kenya
Somalia (Somaliland)	Lesotho	Liberia
Sudan	Liberia	Mozambique
Zimbabwe	Malawi	Niger
	Mozambique	Nigeria
	Niger	Pakistan (Balochistan)
	Nigeria	Somalia (Puntland)
	Pakistan (Balochistan)	Somalia (Somaliland)
	Somalia (Puntland)	Sudan
	Somalia (Somaliland)	Togo
	Sudan	Uganda
	Togo	Yemen
	Uganda	
	Yemen	
	Zimbabwe	

In the EiE sector, regional and NGO documents reference common standards for learning set out by the INEE Minimum Standards. The LSCE MENA framework points to specific skills outlined in the INEE standards. ECW uses the INEE Minimum Standards as part of the benchmark for instructional/learning processes across ECW-supported programs.

There is also evidence that some program approaches deliberately link to regional and national assessments and standards. NGOs that are implementing programs for learning often use national, regional, and international assessments for monitoring and measurement. For example, when FHI360 implements accelerated education programs, SEL, and workforce-preparation programming, they always

include learning outcome assessments like ASER and EGRA/EGMA, based on what is needed and relevant for the planned intervention. Other NGOs with strong technical teams develop new curriculum and assessment tools for education interventions. Save the Children, for instance, has developed and administered HALDO, IDELA, Literacy Boost, and Numeracy Boost assessments as diagnostics for skills, national monitoring, and comparing the effectiveness of interventions, impact evaluation, and program evaluations.

Summary

In conflict- or crisis-affected countries, quality education is a key component of humanitarian response. However, measuring learning in such contexts can be challenging.

Overall, we found that guidance documents, including the SDGs, did not provide robust details on the academic domains that should be measured. Further work by organizations such as the UIS has operationalized some of the global guidance documents for SDG4 and Education 2030, and links them to existing assessment tools (USAID & UIS, 2019). For the six other guidance documents reviewed for this report, the learning domains remain open to interpretation and, as a result, are not used extensively in the monitoring and evaluation of learning outcomes at a national or program level.

Furthermore, none of the guidance documents reviewed included specific guidance on measuring learning outcomes in EiE, such as how to ensure that children affected by conflict or crisis are included in a study sample or what contextual information to collect in EiE settings. Of the 30 measurement tools examined, only three were developed specifically for EiE contexts—Save the Children’s HALDO, the UNRWA MLA, and the EDC OLA. Among the remaining tools, we found the following:

- In crisis contexts, existing assessment systems are not set up to capture populations on the move or respond inadequately to additional strain. Moreover, without a consistent set of goals for what children should know and be able to do, it is difficult to build a functional learning assessment system. This results in a fragmented system with non-comparable data and few or no baseline reference points. In addition, when displaced populations are integrated into national systems, they are not registered as refugees or IDPs. While this may be a positive step in terms of reducing stigma, it makes it impossible to track progress over time and makes it difficult for policy-makers to address these children’s specific needs.
- Relatedly, where parallel systems of education provision exist, fragmentation is amplified. For instance, some refugees attend the host country public schools while others reside in camps where education services are provided by the international community or local NGOs.
- National governments, particularly in countries affected by conflict and crisis, are often unable to collect data and manage assessment systems. Because humanitarian aid provided for education in crisis contexts is so short term, this can limit the implementing agencies’ ability to help strengthen national and local institutional capacity to collect data.

- In some situations, particularly when refugee or displaced populations are not integrated into the local education system, such as the Rohingya refugee crisis in Bangladesh, neither the host country nor the country of origin allows their national curriculum to be used with refugee children (McPherson & Paul, 2019).
- Children and youth who have faced conflict and crisis may bear a cognitive load far greater than that of their peers. This means that conflict- or crisis-affected populations would be disadvantaged in taking standardized tests that have been validated with populations not facing the same difficulties. Traditional test protocols also may not be appropriate, as they could cause additional frustration and stress for students and parents.

All of the tools examined have conducted some type of reliability or validity testing. The degree to which these results are transparent varies. In analyzing the cross-national, regional, and international assessment tools that met our criteria, we found that decoding, reading comprehension, and number sense, which constitute what might be considered the basic literacy and numeracy competencies, were widely captured by the majority of tools. This finding was reflected when we looked at a small sample of national assessments. There are also tools in our sample that assessed more complex literacy and mathematics competencies, and a small number of them were developed specifically for EiE settings. Many of these tools can be adapted, but that requires assistance from experts in both learning assessment and EiE.

The guidance documents, measurement tools, and program approaches mapped in this paper have the potential to increase equity for children in conflict-affected and fragile contexts, but by and large they currently do not. Including EiE contexts in the development and validation of measurement tools and sampling areas of countries affected by conflict or crisis are ways governments and non-government actors can include EiE settings in the measurement of learning outcomes. Researchers and policy-makers can improve equity in EiE settings by collecting contextual data and using that information to explain differences among children. Finally, it is important that data from these measurement efforts are used to start a conversation about the quality of learning opportunities available for children in conflict- or crisis-affected areas, and that the results are acted on.

Bibliography

- ACER Center for Global Education Monitoring. (2015). *Report on the concurrent validity and inter-rater reliability studies of Uwezo*. Washington DC: Results for Development Institute (R4D)
- African Union. (2016a). *Continental education strategy for Africa (CESA 16-25): Indicators manual*. Addis Ababa: African Union.
- African Union. (2016b). *Continental education strategy for Africa (CESA 16-25)*. Addis Ababa: African Union.
- Bεekunko. (2015). *Manuel de formation des enqueteurs*. Bamako, Mali: OMAES
- Belisle, M., Cassity, E., Kacilala, R., Seniloli, M. T., & Taoi, T. (2016). *Pacific Islands Literacy and Numeracy Assessment: Collaboration and innovation in reporting and dissemination*. Bangkok: UNESCO, NEQMAP, ACER.
- Bergin, C. (2017). *Promising practices in refugee education: Synthesis report*. London: Save the Children.
- BRAC & Institute of Informatics and Development (IID). (n.d.). *An assessment of children's basic learning competencies*. Mohakhali, Dhaka: BRAC.
- Byamugisha, A., & Ssenabulya, F. (2005). *The SACMEQ II project in Uganda: A study of the conditions of schooling and the quality of education*. Harare: SACMEQ.
- Chabaditsile, G. K., Galeboe, A. K., & Nkwane, T. D. (2018). *The SACMEQ IV project in Botswana: A study of the conditions of schooling and the quality of education (SACMEQ Policy Research Report No. 4)*. Gaborone, Botswana: Department of Educational Planning and Research Services.
- Chetty, M., Moloji, M. Q., Poliah, R. R., & Tshikororo, J. (2017). *The SACMEQ IV project in South Africa: A study of the conditions of schooling and the quality of education*. Pretoria, South Africa: Department of Basic Education.
- Chinna, U. (2016). *Kingdom of Cambodia Nation Religion King: Grade six national assessment findings in 2016*. Phnom Penh, Kingdom of Cambodia, Ministry of Education, Youth & Sport
- Clarke, M. (2012). *What matters most for student assessment systems: A framework paper*. Washington DC: The International Bank for Reconstruction and Development/The World Bank
- Comings, J. P., Strucker, J., & Bell, B. (2017). Two reading assessments for youth in alternative basic skills and livelihood skills training programs. *Prospects*, 47(3), 229-243.
- Diop, N. M. (2017). *Evaluation internationale PASEC2019*. Dakar: CONFEMEN. Retrieved from <http://www.pasec.confemen.org/evaluation/evaluation-internationale-pasec2019/>
- D'Sa, N., Noble, F., & Pisani, L. (2019). *Collaborative change: Leveraging international IDELA partnerships to benefit young children*. Save the Children.
- Dwardkan, I. (2017). *The SACMEQ IV study in Mauritius: A study of the conditions of schooling and the quality of education*. Gaborone, Botswana: Southern and Eastern Africa Consortium for Monitoring Education Quality.

- Education Cannot Wait. (2017). *Operational model & results framework*. Retrieved from <https://www.educationcannotwait.org/download/results-framework/>
- eOLA. (2013). *Expert review of EDC's Out-of-school Literacy Assessment (OLA)*. Boston: Education Development Center.
- Flotts, M. P., Manzi, J., Jiménez, D., Abarzúa, A., Cayuman, C., & García, M. J. (2015). *Results report TERCE: Learning achievements*. Santiago, Chile: UNESCO Publishing.
- Gay, J., & Cole, M. (1967). *The new mathematics and an old culture: A study of learning among the Kpelle of Liberia*. New York: Holt, Rinehart, and Winston.
- Global Partnership for Education (GPE). (2017). *GPE 2020 results framework*. Retrieved from <https://www.globalpartnership.org/sites/default/files/2018-06-gpe-2020-results-framework-2017.pdf>
- Global Partnership for Education (GPE). (2018). *Methodology sheet for GPE results indicators*. Retrieved at https://www.globalpartnership.org/sites/default/files/methodological_sheet_for_indicator_18.pdf
- Gochyyev, P., Mizunoya, S., & Cardoso, M. (2019). *Validity and reliability of the MICS foundational learning module* (MICS Methodological Papers, No. 9, Data and Analytics Section, Division of Data, Research and Policy). New York: UNICEF.
- Gove, A., & Wetterberg, A. (Eds.). (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI International.
- Hattori H., Cardoso, M., & Ledoux, B. (2017). *Collecting data on foundational learning skills and parental involvement in education* (MICS Methodological Papers, No. 5, Data and Analytics Section, Division of Data, Research and Policy). New York: UNICEF.
- INEE. (2010). *Minimum standards for education: Preparedness, response, recovery*. New York: UNEE Coordinator for Minimum Standards and Network Tools.
- Krupar, A., D'Sa, N., Westrope, C., & Johna, J. F. (2019). Developing a holistic assessment of children's learning in the context of forced displacement: Case study from Dadaab, Kenya. In *Data collection and evidence building to support education in emergencies* (NORRAG Special Issue 02). Geneva, Switzerland: Save the Children.
- Laboratoire de Recherche sur les Transformations Economiques et Sociales (LARTES). (2014). *Baromètre de mesure de la qualité des apprentissages au Sénégal: Livret de l'animateur*. Dakar, Senegal: Jàngandoo.
- LEARNigeria. (2015). *Instruction manual*. Lagos, Nigeria: TEP Center.
- Loizillon, A., Petrowski, N., Britto, P., & Cappa, C. (2017). *Development of the early childhood development index in MICS surveys* (MICS Methodological Papers, No. 6, Data and Analytics Section, Division of Data, Research and Policy). New York: UNICEF.
- Martin, M., & Hastedt, D. (2016). *IEA's LaNA for developing countries*. Retrieved from http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/08/5_IEA-LaNa-for-GAML-meeting_20160511.pdf
- Martin, M. O., & Mullis, I. V., (Eds.). (2015). *PIRLS 2016 assessment framework*. Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement.

- Martin, M. O., Mullis, I. V., & Foy, P. (2016). Assessment design for PIRLS, PIRLS literacy, and ePIRLS in 2016. *PIRLS*, 55-69.
- Martin, M. O., Mullis, I. V., & Hooper, M. (2017). *Methods and procedures in PIRLS 2016*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>
- McPherson, P., & Paul, R. (2019). *Rohingya "lost generation" struggle to study in Bangladesh camps*. Cox's Bazar, Bangladesh: Reuters. Retrieved from <https://www.reuters.com/article/us-myanmar-rohingya-education/rohingya-lost-generation-struggle-to-study-in-bangladesh-camps-idUSKCN1QZ0EA>
- Mendenhall, M. (Ed). (2019). *Data collection and evidence building to support education in emergencies* (NORRAG Special Issue 02). Geneva, Switzerland: Save the Children.
- Mullis, I. V., Martin, M. O., & Sainsbury, M. (2016). PIRLS 2016 reading framework. *PIRLS*, 11-29.
- Mullis, I. V., & Martin, M. O. (Eds.) (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement.
- Nirrengarten, A., Pulfer, G., & Weisenhorn, N. (2018). *INEE Steering Group business case form: Learning outcomes in EIE*.
- Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Oxford, United Kingdom: Blackwell Publishers.
- OECD. (n.d.). *PISA for development technical report*. Paris: OECD Publishing. Retrieved from <https://www.oecd.org/pisa/pisa-for-development/pisaforddevelopment2018technicalreport/>
- OECD. (2016-2019). *PISA for development brief series*. Retrieved from <https://www.oecd.org/pisa/pisa-for-development/pisaforddevelopment-documentation-briefs.htm>
- OECD. (2017). *PISA for development assessment and analytical framework: Reading, mathematics and science, preliminary version*. Paris: OECD Publishing.
- OECD. (2018). *PISA for development assessment and analytical framework: Reading, mathematics and science*. Paris: OECD Publishing.
- Pacific Community. (2016). *2015 Pacific Islands Literacy and Numeracy Assessment (PILNA): Regional report*. Suva, Fiji: Educational Quality Assessment Program.
- Pacific Community. (2018). *PILNA 2018 implementation manual—school coordinator*. Suva, Fiji: Educational Quality Assessment Program.
- Pacific Islands Forum Secretariat & University of the South Pacific. (2018). *Pacific Regional Education Framework (PacREF) 2018-2030: Moving towards Education 2030*. Retrieved at <https://pacificdata.org/data/dataset/pacific-regional-education-framework-pacref-2018-2030-moving-towards-education-2030>
- PAL Network. (n.d.). *Tests and tools*. Retrieved from <https://palnetwork.org/tools/>
- PAL Network. (n.d.). *Report: Refugee children outperforming Ugandans*. Retrieved from <https://palnetwork.org/report-refugee-children-out-performing-ugandans/>

- PASEC. (n.d.). *Evaluation internationale PASEC2014*. Dakar: PASEC, CONFEMEN. Retrieved from <http://www.pasec.confemen.org/evaluation/?annee=PASEC2014>
- PASEC. (2014). *Qualité de l'enseignement fondamental au Mali: Quels enseignements ? Rapport d'évaluation diagnostique au Mali 2011/2012*. Dakar: CONFEMEN.
- PASEC. (2015). *PASEC 2014 education system performance in francophone Sub-Saharan Africa: Competencies and learning factors in primary education*. Dakar: PASAC CONFEMEN.
- PASEC. (2017). *Rapport technique de l'évaluation internationale PASEC2014*. Dakar: PASEC, CONFEMEN.
- Perry, B., Young-Loveridge, J., Dockett, S., & Doig, B. (2008). The development of young children's mathematical understanding. In Forgasz, et al. *Research in mathematics education in Australasia 2004-2007* (pp. 17-40). Rotterdam/Taipei: Brill Sense.
- Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment. *International Journal of Educational Research*, 91, 1-15.
- Platas, L. M., Ketterlin-Gellar, L., Brombacher, A., & Sitabkhan, Y. (2014). *Early Grade Mathematics Assessment (EGMA) toolkit*. Research Triangle Park, NC: RTI International.
- Ramaswami, B., & Wadhwa, W. (2010). *Survey design and precision of ASER estimates*. Retrieved from http://img.asercentre.org/docs/Aser%20survey/Technical%20Papers/precisionofaserestimates_ramaswami_wadhwa.pdf
- RTI International. (2015). *Early Grade Reading Assessment (EGRA) toolkit* (2nd ed.). Washington, DC: USAID.
- RTI International. (2016). *Ghana 2016 national education assessment: Report of findings*. Washington, DC: USAID.
- Save the Children. (n.d.). *Literacy Boost*. Retrieved from <https://www.savethechildren.org/us/what-we-do/global-programs/education/literacy-boost>
- Save the Children. (2012). *Numeracy Boost toolkit: General overview*. Retrieved from Save the Children website at <https://resourcecentre.savethechildren.net/library/numeracy-boost-toolkit-general-overview>
- Save the Children. (2013). *Beyond school walls: A boost for readers*. Retrieved from Save the Children website at <https://resourcecentre.savethechildren.net/library/beyond-school-walls-boost-readers>
- Save the Children. (2017). *International Development and Early Learning Assessment (IDELA) tool*. Retrieved from Save the Children website at <https://resourcecentre.savethechildren.net/library/international-development-and-early-learning-assessment-technical-paper>
- Save the Children. (2018). *Holistic Assessment of Learning and Development Outcomes (HALDO): Administration and adaptation guidance*.
- Saxe, G. (1991). *Culture and cognitive development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- SCALE Ghana. (2016). *Scale Ghana 2016 instructions booklet*. Retrieved from https://palnetwork.org/wp-content/uploads/2016/2016Tools/003_Ghana_InstructionBooklet.pdf
- SEA-PLM. (n.d.) Methodology. Retrieved from http://www.seaplum.org/index.php?option=com_content&view=article&id=26&Itemid=248

- SEA-PLM. (2017). *Assessment framework: Revised draft document incorporating feedback from the Domain Technical Review Teams*. SEA-PLM, UNICEF, ACER. https://research.acer.edu.au/ar_misc/26/
- Shigwedha, A. N., Nakashole, L., Auala, H., Amakutuwa, H., & Ailonga, I. (2017). *The SACMEQ IV project in Namibia: A study of the conditions of schooling and the quality of primary education in Namibia* (SACMEQ Policy Research Report No. 4). Windhoek: Ministry of Education, Arts and Culture.
- Soares, F., Smiley, A., & Lavan, D. (2018). *Education equity indicators for access: Guidance for practitioners in crisis and conflict-affected contexts*. Washington, DC: Education Equity Research Initiative.
- Spink, J. (2018). *Southeast Asia Primary Learning Metrics (SEA-PLM) phase II: Field trial final report*. Retrieved from Australian Council for Education Research website: https://research.acer.edu.au/ar_misc/26
- Treviño, E., & Órdenes, M. (2017). *Exploring commonalities and differences in regional and international assessments* (Information Paper No. 48). Montreal: UNESCO Institute for Statistics.
- UNESCO. (2014). *Comparison of results between the Second and the Third Regional Comparative and Explanatory Studies: SERCE and TERCE 2006-2013*. Santiago, Chile: UNESCO Publishing.
- UNESCO. (2016). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*. Retrieved at <https://iite.unesco.org/publications/education-2030-incheon-declaration-framework-action-towards-inclusive-equitable-quality-education-lifelong-learning/>
- UNESCO, UNICEF, Brookings Institution, & World Bank. (2017). *Overview: Measuring Early Learning Quality and Outcomes (MELQO)*. Retrieved at <https://unesdoc.unesco.org/ark:/48223/pf0000248053>
- UNESCO IBE & UNESCO Institute for Statistics (UIS). (2017). *Monitoring progress towards SDG 4.1: Initial analysis of national assessment frameworks for mathematics*. Geneva and Montreal: UNESCO IBE, UIS. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000259685>
- UNESCO IBE & UIS. (2018). *Monitoring progress towards SDG 4.1: Initial analysis of national assessment frameworks for reading*. Geneva and Montreal: UNESCO IBE, UIS. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000265621>
- UNESCO OREALC. (2016). *Reporte tecnico. Tercer estudio regional comparativo y explicativo*. Santiago, Chile: TERCE.
- UNESCO Institute for Statistics (UIS). (2016a). *Understanding what works in oral reading assessments: Recommendations from donors, implementers and practitioners*. Montreal: UIS.
- UNESCO Institute for Statistics. (2016b). *The cost of not assessing learning outcomes* (UIS Information Paper 26). Montreal: UIS.
- UNESCO Institute for Statistics. (2018a). *Making the case for a learning assessment: Quick guide no. 2*. Montreal: UIS.
- UNESCO Institute for Statistics. (2018b). *SDG4 data book: Global education indicators 2016*. Montreal: UIS.
- UNESCO Institute for Statistics, FHI360, Oxford Policy Management, & REAL. (2018). *Handbook on measuring equity in education*. Montreal: UIS.

- UNHCR. (2018). *Experts agree on important standards and recommendations for refugee statistics*. Retrieved from <https://www.unhcr.org/blogs/experts-finally-agree-on-standards-for-refugee-statistics/>
- UNICEF. (2017). *Reimagining life skills and citizenship education in the Middle East and North Africa: A four-dimensional and systems approach to 21st century skills. Conceptual and Programmatic Framework*. Amman: UNICEF MENA Regional Office.
- UNICEF. (2018). *Breaking all barriers: 8 million children sit end-of-year exams across the Middle East and North Africa*. Amman: UNICEF MENA Regional Office.
- UNRWA. (2018). *Annual operational report 2017: For the reporting period 01 January-31 December 2017*. Amman, Jordan: UNRWA
- USAID. (n.d.). *What is Literacy Boost?* Retrieved from https://earlygradereadingbarometer.org/pages/about_literacy_boost
- USAID & UIS. (2019). *Global proficiency framework: Reading and mathematics*. Montreal: UIS.
- Uwezo. (2017). *Are our children learning? Lessons from Uwezo learning assessments from 2011 to 2015*. Dar es Salaam: Twaweza East Africa.
- Vagh, S. B. (2009). *Evaluating the reliability and validity of the ASER testing tools*. New Delhi: ASER Centre. Retrieved from <http://www.asercentre.org/sampling/precision/reliability/validity/p/180.html>
- Vagh, S. B. (2010). *Validating the ASER testing tools: Comparisons with reading fluency measures and the Read India measures*. Retrieved at <https://pdfs.semanticscholar.org/b760/f26dd600b30d9e728209c45f8cbf2a047322.pdf>
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23(3), 34-41.
- Wagner, D. (2011). *Smaller, quicker, cheaper: Improving learning assessments for developing countries*. Paris: UNESCO-IIEP
- Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, 41, 21-36.
- World Vision International. (2017). *STAR Guidance (School-based Test About Reading): A guide to STAR tool preparation and contextualization*. Retrieved at https://www.wvi.org/sites/default/files/STAR%20Guidance_v1.3.pdf

Annex A. List of Stakeholders Consulted

Diogo Amaro, UNICEF
Roxane Caires, 3EA
Luis Crouch, RTI
Frosse Dabit, UNRWA
Seamus Hegarty, IEA/University of Warwick
Adolfo Gustavo Imhof, UIS
Julius Kinyash, Kakuma Refugee Camp
Patrick Montjourides, NORRAG
Sebastian Meyer, IEA
Antoine Marivin, SEA-PLM
Audrey Nirrengaarten, AfD
Silvia Montoya, UIS
Maria Layhing Siu Munro, ArmandoAid
Clara Pava, Save the Children
Gerhard Pulfer, Porticus
Mary Sugure, EDC
Ramya Vivekanandan, GPE

Annex B. Guidance Documents

Table 15. Global frameworks included in analysis

Framework	Agency	Coverage
SDG 4	UIS is custodian agency for all indicators except 4.2, which is the mandate of UNICEF	Global
Education 2030	UNESCO/UIS	Global
INEE Minimum Standards	INEE	All education in emergencies settings
Agenda 2063 and Continental Strategy for Education in Africa 2016-2025	African Union	55 African countries (northern and sub-Saharan)
Reimagining Life Skills and Citizenship Education in the Middle East and North Africa: Conceptual and Programmatic Framework	UNICEF	MENA region
Pacific Regional Education Framework	Countries of the Pacific Islands Forum	Pacific Island countries
Global Partnership for Education (GPE) Results Framework	GPE	72 developing country partners, around half of which are considered fragile or conflict affected
Education Cannot Wait (ECW) Results Framework	ECW/UNICEF	19 crisis-affected ECW countries

SDG4 GLOBAL FRAMEWORK

In terms of scope, the SDG4 Framework and Education 2030 documents are the widest reaching on this list and are considered the global goal-posts for education, outlining standards and targets to help guide programming at the international, country, and local level. Education 2030 emphasizes that, upon completion of primary and secondary education, all children should have the establish building blocks of basic literacy¹⁸ and numeracy,¹⁹ and achieved learning outcomes as measured against curricula and official standards. SDG4 includes specific monitoring indicators to this end, and specifies assessments that meet the criteria to assess these indicators:

- Indicator 4.1.1 monitors the proportion of children and young people in (a) grades 2/3, (b) at the end of primary, (c) at the end of lower secondary achieving minimum proficiency level in reading and mathematics, by sex.
- Indicator 4.1.2 monitors the proportion of children under 5 years that can identify at least 10 letters of the alphabet, read 4 simple words, and name all numbers from 1 to 10 as measured through household surveys.

¹⁸ “Literacy” is defined as the ability to identify, understand, interpret, create, communicate, and compute using printed and written materials associated with diverse contexts.

¹⁹ Under numeracy, there is general mention of manipulation of numbers, problem-solving, measurement, ratios, and quantities as key life skills.

The SDGs do not provide common definitions or constructs of minimum proficiencies for what should be measured within reading and mathematics. In August 2019, the Global Alliance to Monitor Learning (GAML) met to agree upon minimum proficiency standards for literacy and numeracy.²⁰

INEE MINIMUM STANDARDS

INEE Minimum Standards is the global guidance document for education in emergencies settings, and outlines that functional literacy, numeracy, and essential life skills to attain a life with dignity and participate meaningfully in the community are core competencies. The framework includes guidance on how to conduct assessments, in particular community assessments, but no particular tools or learning domains are mentioned.

REGIONAL GUIDANCE DOCUMENTS

Regional guidance documents like the Continental Strategy for Education in Africa 2016-2025 (CSEA), Reimagining Life Skills and Citizenship Education in the MENA: Conceptual and Programmatic Framework (LSCE MENA), and the Pacific Regional Education Framework (PacRef) highlight the core competencies of reading, writing, and math, but do not include more specific sub-domains. Some do provide general guidance on measuring learning. CSEA and LSCE MENA set out targets for monitoring progress on learning that build directly on SDG indicators, including the main SDG indicator measuring learning: the proportion of children in grade 3, at the end of primary, and at the end of lower secondary achieving at least a minimum proficiency level in reading and math. CSEA mentions the sources of this indicator generally as national and cross-national assessments, and the Life Skills and Citizenship Education in MENA framework highlights MICS or other household surveys, TIMSS, and PISA. The Pacific Regional Education Framework, on the other hand, does not identify numeric targets for region-wide gains in learning, as it assumes each national education system sets their own learning targets. Instead, it sets out broad-based goals of establishing upward trends in learning.

There is some mention of conflict and education in emergencies in the regional guidance documents reviewed. CSEA's strategic objectives include promotion of peace education and conflict prevention, and mention the importance of education in conflict zones, especially for girls. The LSCE MENA document also emphasizes peace education, outlining that "schools and classrooms can either become the space for social cohesion or for further entrenching inequity and the structural roots of conflict." It also strongly emphasizes that higher order skills like problem-solving, decision-making, persistence, creative thinking, etc., are essential in relation to conflict and crisis contexts, referencing the Learning Metrics Task Force and the OECD PISA Analytical Framework as key documents identifying skills for lifelong learning.

²⁰ <http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/GAML6-REF-16-GLOBAL-PROFICIENCY-FRAMEWORK.pdf>

PROGRAM RESULTS FRAMEWORKS

The monitoring and results frameworks for agencies like the Global Partnership for Education (GPE) and Education Cannot Wait (ECW) target the focus countries of their organizations and test the effectiveness of their programmatic models. While literacy and numeracy skills are mentioned generally, indicators included in these frameworks do not specify domains or sub-domains. Indicators used to measure progress on the frameworks examine minimum standards of learning using data from existing national or international learning assessments.

Annex C. Measurement Tools

TOOLS FOR SYSTEM MONITORING

Table 16. System-monitoring measurement tools included in mapping exercise with internationally/regionally-comparable results

Name of Assessment	Acronym	Grade or Age	Subject(s) Assessed	EiE Country Contexts	Administering Agency
Multiple indicator cluster survey: early childhood development index	MICS-ECDI	Ages 3-4	Literacy-numeracy, physical development, social-emotional skills	Bangladesh, Benin, Cameroon, Chad, CAR, Congo/DRC, Cote d'Ivoire, El Salvador, Gambia, Ghana, Guinea, Guinea-Bissau, Honduras, Indonesia, Iraq, Kenya, Kyrgyzstan, Lebanon, Lesotho, Malawi, Mali, Mauritania, Nigeria, Pakistan, Palestine, Senegal, Sierra Leone, Somalia, South Sudan, Sudan, Togo, Tonga, Tunisia, Ukraine, Zimbabwe	UNICEF
Multiple indicator cluster survey: foundational learning skills module*	MICS-FLS	7- to 14-year-olds	Reading, numeracy skills	Bangladesh, Benin, Chad, CAR, Congo/DRC, El Salvador, Gambia, Ghana, Guinea-Bissau, Honduras, Iraq, Kyrgyzstan, Lebanon, Lesotho, Malawi, Pakistan, Palestine, Sierra Leone, Sudan, Togo, Tonga, Tunisia, Zimbabwe	UNICEF
Programme d'analyse des systèmes éducatifs de la CONFEMEN	PASEC	Grade 2, grade 6	Reading, mathematics	Benin, Burkina Faso, Burundi, Cameroon, Chad, Congo/DRC, Cote d'Ivoire, Guinea, Mali, Niger, Senegal, Togo	CONFEMEN
Latin American Laboratory for Assessment of the Quality of Education	LLECE	Grade 3, grade 6	Reading and writing, mathematics, science (grade 6)	Colombia, El Salvador, Guatemala, Honduras, Nicaragua, Venezuela	UNESCO-Regional Bureau for Education in Latin America and the Caribbean
Progress in International Reading Literacy Study	PIRLS	Grade 4	Reading	Colombia, Turkey	International Association for the Evaluation of Educational Achievement (IEA)

Literacy and numeracy assessment	LaNA	Grade 4 (or 5, 6)	Basic reading, basic mathematics	Haiti, Nigeria, Pakistan (Punjab)	IEA
Pacific Islands Literacy and Numeracy Assessment	PILNA	Grade 4, grade 6	Literacy, numeracy	Papua New Guinea, Solomon Islands, Tonga	Education Quality and Assessment Program, Pacific Community
Trends in International Mathematics and Science Study	TIMSS	Grade 4, grade 8	Mathematics, science	Colombia, El Salvador, Ghana, Indonesia, Jordan, Lebanon, Pakistan, Palestine, Philippines, Syria, Tunisia, Yemen	IEA
Southeast Asia Primary Learning Metric	SEA-PLM	Grade 5	Reading and writing, mathematics, global citizenship	A set of countries will undertake trial including the Philippines, Myanmar, and Cambodia	Southeast Asian Ministers of Education Association/ UNICEF East Asia Pacific Regional Office
Southern and Eastern Africa Consortium for Monitoring Educational Quality	SEACMEQ	Grade 6	Reading, mathematics, HIV/AIDS knowledge	Angola, Kenya, Lesotho, Malawi, Mozambique, Uganda, Zambia, Zimbabwe	SEACMEQ
Programme for International Student Assessment for Development	PISA-D	15-year-olds (includes OOSC)	Reading, mathematics, science	School-based component: Cambodia, Guatemala, Honduras, Senegal, Zambia Out-of-school component: Guatemala, Honduras, Senegal, Zambia	PISA

Multiple indicator cluster survey: Early childhood development index (MICS-ECDI)

AGES 3-4 | LITERACY-NUMERACY, PHYSICAL DEVELOPMENT, SOCIAL-EMOTIONAL SKILLS

The ECDI is one of the first population-based measures of early childhood development available at an internationally comparative level. It is included in Multiple Indicator Cluster Surveys (MICS), the UNICEF-supported international household survey. It includes items in four early developmental domains: language/cognitive, physical, socio-emotional, and approaches to learning. Due to its brevity and simplicity (10 items, with binary responses), the ECDI is easy to administer, calculate, and interpret.

Multiple indicator cluster survey: Foundational learning skills module* (MICS-FLS)

AGES 7-14 | LITERACY, NUMERACY

The FLS instrument of reading and mathematics skills, newly developed starting in 2014, is a self-contained module within the broader MICS household survey. The module assesses basic reading and numeracy skills at a grade 2 level and is administered to all children ages 7-14, regardless of whether they attend school or what grade they attend. The module has three sections: a short interview to the child, a reading task, and a numeracy task. Responses are

oral and tasks are not timed. The module is designed to produce learning outcome data that can be compared across multiple languages and countries.

Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC)
GRADES 2 & 6 | LITERACY, NUMERACY

PASEC is a regional assessment tool for Francophone countries in West Africa and Asia, conducted by the Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie (CONFEMEN). It provides information about the performance of students at the beginning and end of primary school in member countries to understand the effectiveness and equity of national education systems.

Latin American Laboratory for Assessment of the Quality of Education (LLECE)
GRADES 3 & 6 | LITERACY, WRITING, MATHEMATICS, SCIENCE

LLECE's PERCE, SERCE, TERCE, and ERCE are large-scale studies of learning achievement coordinated by the Latin American Laboratory for Assessment of the Quality of Education (LLECE), based in the UNESCO Regional Bureau for Education in Latin America and the Caribbean. From the initial focus on reading and mathematics, the approach has grown to include writing and the natural sciences. LLECE assessments aim to identify the factors associated with different levels of achievement, such as the socioeconomic context, family life and personal issues, educational policies, and school processes. The third cycle, TERCE, was implemented in 2013. A new cycle of the Regional Comparative and Explanatory Study is underway in 2019 (called ERCE) and will include a new section on socio-emotional skills measurement.

Progress in International Reading Literacy Study (PIRLS)
GRADE 4 | LITERACY

The PIRLS is an international comparative assessment that measures student learning in reading. Since 2001, PIRLS has been administered every 5 years. PIRLS documents worldwide trends in the reading knowledge of 4th graders, as well as school and teacher practices related to instruction. Fourth-grade students complete a reading assessment and questionnaire that addresses students' attitudes towards reading and their reading habits. Questionnaires are given to teachers and school principals to gather information about school experiences.

Literacy and Numeracy Assessment (LANA)
GRADES 4, 5, or 6 | LITERACY, NUMERACY

A short basic assessment at the end of primary school begun in 2015, LaNA is intended for countries where the IEA's Trends in International Mathematics and Science Study and Progress in International Reading Literacy Study (TIMSS and PIRLS) may be too difficult to implement. For developing education systems that participate in LaNA, the study can serve as a stepping stone for participation in future IEA TIMSS and PIRLS cycles.

Pacific Islands Literacy and Numeracy Assessment (PILNA)
GRADES 4 & 6 | LITERACY, NUMERACY

PILNA is a measurement of regional standards based on a common scale; it is a regional collaborative model that is highly consensual among the participating countries and directly linked to their commitment to the SDGs. It was

administered in 2012 as a one-time snapshot of literacy and numeracy levels in the region; education ministers requested a second administration in 2015.

Trends in International Mathematics and Science Study (TIMSS)

GRADES 4 & 8 | MATHEMATICS, SCIENCE

TIMSS is a research study conducted every 4 years by the International Association for the Evaluation of Educational Achievement (IEA) that measures mathematics and science achievement at grades 4 and 8. In addition, it collects information on curriculum and curriculum implementation, instructional practices, and school resources. For countries where students are still developing fundamental mathematics skills, TIMSS Numeracy assessment (designed to be administered at grades 4, 5, or 6) concentrates on measuring numeracy learning outcomes, including fundamental mathematical knowledge, procedures, and problem-solving strategies. TIMSS Advanced measures trends in advanced mathematics and physics for final-year secondary school students.

Southeast Asia Primary Learning Metric (SEA-PLM)

GRADE 5 | LITERACY, WRITING, NUMERACY, GLOBAL CITIZENSHIP

SEA-PLM is a regionally contextualized primary school learning assessment including common metrics for reading, writing (across languages and scripts), math, and global citizenship, developed by analyzing curricula and needs of Member Countries. It includes background questionnaires on contextual information. It aims to document changes by repeating cycles of assessment every four years. Data have been collected in a nationally representative sample of schools in each country.

Southern and Eastern Africa Consortium for Monitoring Educational Quality (SEACMEQ)

GRADE 6 | LITERACY, NUMERACY, HIV/AIDS KNOWLEDGE

SEACMEQ is a regional consortium of 16 education ministries in Southern and Eastern Africa. It conducts large-scale, cross-national research studies assessing the conditions of schooling, performance levels of learners and teachers in literacy and numeracy, as well as levels of basic health knowledge.

Programme for International Student Assessment for Development (PISA-D)

AGE 15 | LITERACY, NUMERACY, SCIENCE

PISA tests the skills and knowledge of 15-year-old students from a sample of randomly selected schools every three years. Students are tested in reading, mathematics, and science, as well as to what extent students can apply their knowledge to real-life situations. PISA-D attempts to increase the resolution of the PISA tests at the lower end of the student distribution, capture a wider range of social and economic contexts, incorporate an assessment of out-of-school 14- to 16-year-old youth, and adds support for building the capacity of participating countries to implement international large-scale assessments and use assessment results to support evidence-based policy-making.

MULTI-PURPOSE TOOLS

Table 17. Multi-purpose measurement tools included in mapping exercise where results cannot be compared internationally

Name of Assessment	Acronym	Grade or Age	Subject Assessed	Context	Administering Agency
Measure of Child Development and Early Learning	MELQO MODEL	Pre-primary	Pre-literacy, pre-numeracy, SEL, executive functioning	10 countries: Bangladesh, Ethiopia, Indonesia, Kenya, Kyrgyzstan, Lesotho, Liberia, Nicaragua, Pakistan, Sudan	World Bank, various agencies
International Development and Early Learning Assessment	IDELA	3.5-6 years	Early reading and math, socio-emotional skills, motor development	Afghanistan, Bangladesh, Cameroon, Colombia, Ethiopia, Ghana, Guatemala, Haiti, Honduras, Indonesia, Jordan, Kenya, Lebanon, Liberia, Malawi, Mali, Mozambique, Myanmar, Nicaragua, Niger, Palestine, Papua New Guinea, Philippines, Senegal, Sierra Leone, Solomon Islands, Somalia, Uganda, Zambia, Zimbabwe	Save the Children
Holistic Assessment of Learning and Development Outcomes	HALDO	4-12 years	Literacy, numeracy, SEL, executive functioning	Uganda, Kenya	Save the Children
Early Grade Reading Assessment	EGRA	Grades 1-3	Basic literacy	Afghanistan, Angola, Bangladesh, Benin, Burundi, Cambodia, Chad, DRC, Ethiopia, Ghana, Guatemala, Haiti, Honduras, Indonesia, Iraq, Jordan, Kenya, Kyrgyzstan, Liberia, Malawi, Mali, Nicaragua, Niger, Nigeria, Pakistan, Palestine, Papua New Guinea, Philippines, Senegal, Sierra Leone, Somalia, South Sudan, Sudan, Syria, Tonga, Uganda, Yemen, Zambia, Zimbabwe	Varies
Early Grade Mathematics Assessment	EGMA	grades 1-3	basic mathematics	Afghanistan, Congo/DRC, Ethiopia, Ghana, Iraq, Jordan, Kenya, Liberia, Malawi, Mali, Mozambique, Nicaragua, Nigeria, Philippines, Sierra Leone, Somalia, Zambia, Zimbabwe	Varies

Citizen-Led Assessments					
ASER (“Impact” in Hindustani)	ASER	5-16 years	Basic reading, basic numeracy	India/Pakistan	Pratham and ASER Centre
UWEZO (“Capability” in Kiswahili)	UWEZO	6-16 years	Basic reading (English, Kiswahili in Kenya/Tanzania, and local languages in Uganda), basic numeracy	Kenya, Tanzania, Uganda	Twaweza
Let’s Engage, Assess & Report Nigeria	Learn Nigeria	5-15 years	Basic reading, basic numeracy	Nigeria	TEP Centre
IID/BRAC Survey	IID/BRAC Survey	5-15 years	Basic reading, basic numeracy	Bangladesh	Institute of Informatics and Development (IID)
Strengthening Citizen Action for Learning and Empowerment Ghana	SCALE Ghana	6-15 years	Basic reading, basic numeracy	Ghana	Choice Ghana
Todos Pelas Crianças em Moçambique (“All for the Children of Mozambique”)	TPC Mozambique	7-16 years	Basic reading, basic numeracy	Mozambique	Facilidade – Institute for Citizenship and Sustainable Development (ICSD)
Beekunko (“The Concern of Everyone” in Bamanakan)	Beekunko	6-14 years	Basic reading, basic arithmetic	Mali	L’oeuvre Malienne d’Aide a l’Enfance du Sahel (OMAES)
Jàngandoo (“Learning Together” in Wolof)	Jàngandoo	6-18 years	Basic reading, basic numeracy, culture tasks	Senegal	Laboratoire de Recherche sur les Transformations Economiques et Sociales (LARTES)

Measure of Child Development and Early Learning (MELQO MODEL)

AGES 3-8 | PRE-LITERACY, PRE-NUMERACY, SOCIO-EMOTIONAL SKILLS, EXECUTIVE FUNCTIONING

MELQO began in 2014 in anticipation of global emphasis on early childhood development. Tools are intended to generate relevant, usable data to guide policies and programs to improve pre-primary education for children 3 to 8 years old. The MELQO MODEL includes a set of core math, literacy, and executive function items, which can be seen as the starting point for national-level assessment of child development and learning. The framework is intended to be further adapted to align with national priorities and goals.

International Development and Early Learning Assessment (IDELA)

AGES 3.5-6 | LITERACY, NUMERACY, SOCIO-EMOTIONAL SKILLS, MOTOR DEVELOPMENT

IDELA is a tool developed by Save the Children to measure skills that allow children to successfully transition into primary school classrooms based on global curricula and standards. IDELA covers four developmental domains: motor development, emergent literacy, early numeracy, and social-emotional skills. The tool is designed to be easy to use, easily translated, and administered in low-resource settings. It can be used for program evaluation, randomized control trials, comparison of different ECCD interventions, school readiness assessments, and national monitoring of ECCD programs. Each assessment takes about 35 minutes per child.

Holistic Assessment of Learning and Development Outcomes (HALDO)

AGES 4-12 | LITERACY, NUMERACY, SOCIO-EMOTIONAL SKILLS, EXECUTIVE FUNCTIONING

HALDO is a rapid assessment tool developed and piloted in 2018 by Save the Children that measures literacy, numeracy, social and emotional learning, and executive functioning in displaced populations. The assessment focuses on a wider age range than other available assessments to account for varied skills in emergencies. Because the assessment is designed for rapid deployment with minimally trained assessors in the immediate onset of displacement, it is not a comprehensive measure of each domain, but it uses dynamic scoring to assess children's skill levels from emergent to advanced. HALDO takes 30 to 40 minutes to complete.

Early Grade Reading Assessment (EGRA)

GRADES 1-3 | LITERACY

EGRA is designed to orally assess the most basic foundation skills for literacy acquisition in early grades. Test components are based on recommendations made by an international panel of reading and testing experts and include timed, 1-minute assessments of letter naming, nonsense and familiar words, and paragraph reading. Additional (untimed) segments include comprehension, relationship to print, and dictation. It was designed as an inexpensive and simple diagnostic of individual student progress in reading. Initially designed as a sample-based "system diagnostic" measure, over time it has been used to generate baseline data; guide the design of instructional programs, program evaluation, cost-effectiveness evaluations; and develop reading indicators and benchmarks. Although EGRA is standardized in its research foundations and underlying principles, it is designed to reflect and respond to local conditions, and comparability across countries and languages is difficult.

Early Grade Mathematics Assessment (EGMA)

GRADES 1-3 | NUMERACY

Building on the success of EGRA, EGMA is an orally administered assessment of the core mathematical competencies taught in primary schools. Because the EGMA is designed for the early grades, which is when children are just beginning to learn how to read, the oral administration does not confound a child's ability to read or write with a child's ability to do mathematics. EGMA is meant to be locally adapted to fit the needs of the local context. For the most part, the EGMA has been used to (1) determine how students in a country are performing overall compared to its stated curriculum, and (2) examine the effectiveness of specific curricula, interventions, or teacher-training programs.

Citizen-led Assessments (ASER, UWEZO, Learn Nigeria, BRAC, SCALE Ghana, TPC Mozambique, Beekunko, Jàngandoo)

AGES 5-18 | LITERACY, NUMERACY

Citizen-led assessments (CLAs) aim to provide reliable estimates of children’s schooling status and basic learning levels. CLAs are household based rather than school based, which allows all children to be included—those who have never been to school or have dropped out, as well as those who are in government schools, private schools, and religious schools. Tests are orally administered and take 10 minutes. CLAs stemmed initially from the ASER tool in India, and have since expanded to countries in varying regions, with each country sharing knowledge with the next. Most CLAs are country specific.

Annex D. Program Approaches

Table 18. Program approaches included in the mapping exercise

Program Approach	Agency	Grade / Age	Context	Subject
Literacy Boost	Save the Children	Grades 1-3	Afghanistan, Bangladesh, Burundi, Cambodia, Colombia, Congo/DRC, Cote d'Ivoire, El Salvador, Ethiopia, Ghana, Guatemala, Haiti, Indonesia, Kenya, Liberia, Malawi, Mali, Mozambique, Niger, Pakistan, Papua New Guinea, Philippines, Senegal, Uganda, Zambia, Zimbabwe	literacy
Numeracy Boost	Save the Children	Grade 2	Bangladesh, El Salvador, Ethiopia, Jordan, Malawi, Pakistan	numeracy
Unlock Literacy - STAR	World Vision	Grade 3	<i>Full country list not available</i>	literacy
Education in Emergencies - MLA	UNRWA	Grade 4, grade 8	Jordan, Lebanon, Palestine, Syria	Arabic, mathematics, science
Read Right Now! - OLA	Education Development Center	Older youth	Ethiopia, Liberia, Mali	literacy

SAVE THE CHILDREN – LITERACY BOOST

Literacy Boost differs from other literacy programs in that it places equal emphasis on literacy acquisition in formal schooling and in communities, and also does not require a change to the national curriculum. Using its adaptive model, teachers gain a better understanding of how children learn to read and write and teach their existing curriculum more effectively. By focusing on core skills, Literacy Boost also helps children learn to read independently, using activities that promote reading motivation and enjoyment. Teachers, students, parents, and communities are thus engaged to help enhance the skills required for independent reading, while fostering growth in children's vocabulary, building their confidence in expression, and expanding their background knowledge that they bring to every reading task.

Literacy Boost has four core components:

- Student Assessments | To identify gaps and measure improvements in the five core reading skills and writing, helping teachers understand the level of learning and tailor their lessons accordingly. Assessment data is shared with government at local and national levels to help inform advocacy and policy change.

- Teacher Training | To incorporate skill-building into their regular curriculum with an emphasis on core reading skills and writing. Ongoing assessments of student progress are used to guide teacher focus.
- Mobilizing communities | Quality learning environments outside of school help children to grow their skills and to foster their love of reading
- Learning Materials | Enhancing the literacy environment ensures the availability of age-appropriate reading materials in local languages (e.g., Book Banks)

In 2017, Save the Children piloted the Learning and Well-being (LWIE) program in Egypt and South Sudan, which adheres to the same pillars of Literacy Boost but with a stronger focus on well-being by measuring children’s well-being as it relates to their learning outcomes, ensuring that teachers have the skills and knowledge to promote social and emotional learning in the classrooms, and engaging the local community in activities that promote literacy and well-being outside of school. The initiative is aimed at determining and using more effective teaching practices with refugee children.

ASSESSMENT TOOL

The Literacy Boost assessment measures six core skills of literacy:

1. Alphabet knowledge – Recognition of the letters of the alphabet
2. Phonemic awareness – The ability to recognize and manipulate the sound units that make up words
3. Fluency – The ability to read accurately, quickly, and with the correct intonation
4. Vocabulary – The comprehension of a sufficient number of words to be able to understand text
5. Comprehension – The process of understanding meaning from written language
6. Writing – The ability to use a set of symbols to represent speech sounds in print and communicate meaning

In its approach to measuring reading skills, Literacy Boost recognizes that young readers possess a spectrum of different abilities, which is why the tool assesses foundational skills, such as concepts about print and letter identification, as well as higher order skills like reading fluency and comprehension. Measuring multiple levels of reading comprehension is important because fluency does not necessarily signal comprehension, and only increasing the speed at which children read is unlikely to improve their overall performance.

In addition to the Literacy Boost tool, Save the Children developed the International Development and Learning Assessment (IDELA) tool, which provides programs and ministries ongoing data on the status of children ages 3.5 to 6. IDELA incorporates measures of emergent language and literacy, early numeracy and problem solving, motor, and socioemotional skills, as well as approaches to learning and executive function (short-term memory and inhibitory control).

Both the Literacy Boost assessment and IDELA use a unique continuous scoring system that allows for a more nuanced perspective on learning and development than is possible if items are simply scored as correct or incorrect. For example, an IDELA measure of expressive vocabulary—a precursor skill to vocabulary measures of oral reading assessments—can be taken by asking a child age 3 to 6 years to name things to eat that can be bought in the market. The number of different items the child names is counted to offer a continuous score across the age range of 3 to 6 years.

The breadth of skills measured by both assessments supplies broader variation for equity analyses and offers more flexibility for interpretation and reporting. In addition, country teams may add additional lower or higher order sub-tests or administer test components in multiple languages, ensuring that the tools are fit for both purpose and context. This, in combination with Save the Children’s commitment to tracking individual pupils rather than sampling

cross-sections of children, aligns with wider programmatic goals of identifying and responding to the unique needs of disadvantaged groups and closing gaps in learning achievement.

SAVE THE CHILDREN – NUMERACY BOOST

Save the Children’s Numeracy Boost, a companion model to the well-established Literacy Boost program, supports students, teachers, parents, and communities in developing the math skills of children in the early primary grades, both inside and outside of the classroom. Numeracy Boost is designed to steer teachers away from methods that rely on memorization, repetition, and student workbooks, instead emphasizing learning through interactive activities and games that offer children the opportunity to explain and explore their reasoning. The model, however, does not require a change to the national curriculum and is intended for adaptation to the national and local context.

Numeracy Boost has three core components: student assessment, teacher training, and community action. First, students’ skills in math are assessed, then teachers are trained to enhance their teaching methodologies and skills to help develop their students’ math knowledge and understanding. Finally, families and communities participate in activities like “math days” and “math camps” with their kids. Kids are introduced to fun math projects—like cooking, going to the market, harvesting—to show how math is an important part of everyday life.

Numeracy Boost was first piloted in Malawi and Bangladesh in 2012 and has since been implemented in full by country offices in Pakistan, Ethiopia, and Egypt. Although Numeracy Boost does not have explicit guidance for humanitarian contexts, the toolkits can be used and adapted to any context, as evidenced by its use in emergency settings in Jordan and Thailand. In Thailand, Save the Children supports refugee education in camps along the border with Myanmar, and it operate child-friendly clubs for Jordanian and Syrian children in Jordan.

ASSESSMENT TOOL

The Numeracy Boost assessment gauges children’s knowledge and skills in three domains—number of operations, geometry, and measurement—through various sub-tests. The assessment also includes a Home Numeracy Background component, which asks about the students’ use of exposure to math outside of school. Assessment data is shared with local and national governments to help inform policy change.

Numeracy Boost focuses on three core domains of mathematics:

1. Number of operations – Understanding how to represent numbers and being able to compute operations accurately
2. Geometry – The identification of shapes, their properties, understanding space, and the relations of points, lines, and angles
3. Measurement – The ability to describe and compare measurable attributes including length, weight, time, and money

WORLD VISION – UNLOCK LITERACY

Since 2012, World Vision has piloted education projects with a focus on building foundational learning and core reading skills for children in the first few years of primary school, rather than simply supporting school attendance. The Unlock Literacy project model was developed following the organization’s experience partnering with Save the Children to implement their Literacy Boost model.

Unlock Literacy promotes five core reading skills: letter knowledge, sounding out words, reading fluency, vocabulary, and comprehension. The core components of the model are:

- Reading Assessments | Baseline and endline reading assessments using World Vision’s STAR tool and mobile tablets measure children’s reading levels, evaluate their learning needs, and help school and ministries of education track students’ progress
- Teacher Training | Teachers learn to incorporate the five core reading skills into their curricula and receive ongoing teacher coaching. Teachers learn to create a print-rich environment in their classrooms and ensure that children remain motivated while learning to read
- Community Action | Mobilizing parents and communities to support children as they learn to read through fun out-of-school reading camps. Communities are also empowered through World Vision’s Citizen Voice and Action model to hold key stakeholders accountable for the delivery of quality educational services.
- Teaching and Learning Materials | Creation of locally relevant and grade-appropriate reading materials. These steps address the ways children learn to read, how teachers and parents can help, and create a culture of reading in school, at home, and in the community.

The model closely matches World Vision’s philosophy of improving children’s education in partnership with government systems, schools, families, and communities. In contrast to typical education projects that focus exclusively on the teacher/school side of the equation, this approach brings a holistic approach that works with teachers, parents, and community members to create a supportive literacy environment that can sustain reading improvements for children, both in and out of school.

ASSESSMENT TOOL

The World Vision School-based Test About Reading (STAR) provides critical information about children’s foundational reading ability. The tool measures a level of minimum proficiency in reading comprehension for grade 3 students attending a structured learning environment (including formal and non-formal school). Components of the assessment include:

1. Letter knowledge – Providing name and/or sound of letters
2. Most used words – Reading words drawn from a list of common words used in grade 3 textbook
3. Decodable words – Stating letter sounds by reading nonsense words
4. Story reading – Reading a grade 3 text with accuracy and at a reasonable rate of speed
5. Reading comprehension – Responding correctly to fact retrieval questions about the story read

Background questions are also asked of students related to sex, home literacy environment, and school environment.

EDUCATION DEVELOPMENT CENTER – READ RIGHT NOW!

The Education Development Center’s (EDC) Read Right Now! (RRN) literacy approach is specifically designed for challenging, resource-lean contexts. The program targets educators with limited training and professional development opportunity, communities where reading and writing are not widespread, systems where classrooms are crowded and materials are sparse and unevenly distributed, and environments where multiple languages are spoken. Settings have included remote, fragile, and high-conflict contexts; public and community-run schools; madrassas; rural one-classroom schools; alternative basic education programs; multi-lingual classrooms; and nomadic schools.

RRN simultaneously addresses oral language development, explicit literacy skills, and authentic reading and writing. The instructional component can be tailored to meet the needs of any context and applied in early grade or low-literacy youth and adult learning classrooms. RRN has four components:

- Enhancing teacher capacity through training and coaching
- Providing print and digital resources for learners and educators
- Engaging families and communities as champions of and supports for literacy
- Ensuring that policies and systems for literacy are coherent and evidence based

ASSESSMENT TOOL

To help teachers and administrators evaluate the literacy level of their students instantly, ECD developed an electronic version of EGRA. “eEGRA” provides a detailed profile of children’s reading ability and allows educators to track literacy improvement over time. eEGRA solves drawback of the paper-based EGRA that limits its usefulness for improving instruction at the classroom level:

1. Turnaround time between assessment and results can take up to six months
2. Results are rarely available below the national or district level

eEGRA was field-tested in the Philippines in July 2010, demonstrating that eEGRA scores learners as accurately as paper-based EGRA and ensuring that it did not introduce any artifacts. ECD continues to refine the tool with new features and improved performance. eEGRA is free to use and open for adaptation.

ECD also developed the Out-of-School Literacy Assessment (OLA) to measure literacy skills of other youth and young adults, particularly those who are living in extreme poverty or post-conflict environments and those with minimal literacy acquisition. Similar to existing child-, youth-, and adult-literacy assessment tools, OLA orally assesses foundational reading skills, including:

- Functional reading (such as food labels and instructions on medicine bottles)
- Letter naming
- Letter sounds/syllables
- Word recognition and decoding
- Oral reading and comprehension
- Silent reading comprehension

In developing the components and individual literacy items, ECD selected tests that could be accurately and reliably scored by assessors who were not trained reading specialists and who may have had little experience giving standardized tests. In addition, a demographic questionnaire includes questions on urban/rural origin, mother tongue, language utilization at home and work, exposure to print outside of classrooms, income, and literacy goals.

OLA can be used for formative and summative purposes, but tests only a sample of youth within programs rather than whole program populations. Sample size is determined by using standard statistical methods.

EDC convened a panel of experts of international assessment, psychometrics, adult literacy, and second-language acquisition in 2013 to review the pilot program of OLA in Liberia. Following the panel's recommendations, EDC strengthened the background questionnaire, revised test items, changed the order in which the sub-tests were given, and shortened some sections with no sacrifice in reliability or validity, based on comparative-item re-analyses of previously collected data.

UNRWA – EDUCATION IN EMERGENCIES

UNRWA is a United Nations agency mandated to provide assistance to Palestine refugees across its five fields of operation—Jordan, Lebanon, Syria, the West Bank, and Gaza. In these areas of crisis, UNRWA strengthens education delivery by adopting a multi-stranded approach that aims to meet the needs of children, teachers, and parents. The Education in Emergencies program consists of 3 components:

- **Ensuring a safe and secure learning environment** that supports the physical and emotional well-being of children and youth. UNRWA puts additional school counselors in place to provide individual and group counseling, provides regular recreational activities, and develops safety and security training modules for education staff and students
- **Stimulating relevant and quality teaching and learning**, including alternative methods of teaching and learning in times of crisis. The UNRWA Department of Education has developed a self-learning program to facilitate the learning of basic skills and core subject concepts in Arabic, mathematics, English, and science. The EiE program also helps provide learning support to help children make up for any lost school time through catch-up classes and summer learning programs.
- **Ensuring parental, community, and student engagement** to support the quality and continuation of education in emergencies. Parents are engaged in supporting their children's education through regular parents' awareness sessions on a variety of topics, including psychosocial support, safety and security, and the self-learning program, and elected student School Parliaments train children on safety and security skills.

There is some evidence that Palestinian children in UNRWA education programs outperform Palestinian children in government-run schools.²¹

ASSESSMENT TOOL

Monitoring of Learning Achievement (MLA) tests are provided to grade 4 and grade 8 students. The MLA test generates information about student performance levels (i.e., in relation to the expected performance at grade

²¹ <http://www.jmcc.org/news.aspx?id=3111>

level); about student learning skills and competencies (i.e., if they are able to reason and apply knowledge or if their competencies are limited to knowledge recall); and about the way in which a subject is taught with regard to its content domains.

Tests are administered to a randomly selected sample of grade 4 and grade 8 students. Both Arabic and math tests comprise two forms, with contain items from the common curriculum across the four fields of operation.

In its design, the MLA survey considered:

1. International best practices in test development, administration and analysis;
2. Differences between host country curricula;
3. The cost and feasibility of implementing a large-scale achievement survey, taking into account the different conditions under which UNRWA operates; and
4. The need for data to lead to analysis which will support decision-making, as well as the identification of strategies, which will help improve student achievement levels.

Since a redesign in 2012, results are no longer given as a mean score, but instead in relation to performance levels—two levels at or above what a student should be able to achieve and two levels below. The ability of children to think critically is also assessed through questions that require higher-order thinking skills in answering. Through questionnaires, the MLA also looks at classroom practices, school environments, and overall equity of student learning outcomes.

Annex E. Descriptions of Domains and Sub-Domains

Figure 10. Content descriptors for reading (UNESCO-IBE and UIS, 2017)

READING COMPETENCY		Decoding, reading comprehension
Decoding	Alphabetic principle (grapheme-phoneme and diagraph-phoneme correspondence, upper and lower case correspondence, recognize the alphabetical order in dictionaries or encyclopaedias); Precision (words, sentences, texts, pronunciation, sight read/direct recognition); Fluency (volume, speed, expressiveness and tone)	
Reading comprehension	Identify (different types of texts, parts of a text, connections/signs/symbols/time/space indicator, parts and types of a sentence, abbreviations/contractions/compound words, purpose for reading); Retrieve (meaning of words, adjacent meaning of a word in a sentence, synonyms of paraphrased terms in a text, information, meaning of a text); Interpret (make inferences, generate conclusions make generalizations, identify evidence for an interpretation in a text, summarize main ideas, apply information from a text to a new context, compare and contrast information, types of texts, figurative language, paratextual information, anticipate the content of a text); Reflect (give an opinion, give reasons for an opinion, generate questions from a text, generate explanations for the information in the text based on personal experiences, assess information from someone else's perspective, about the author, changes made to a text, make critical assessments, combine ideas from different sources to create an opinion or judgment); Metacognition (goal setting and strategy planning, monitoring, evaluation)	
LINGUISTIC COMPETENCY		Listening, speaking, vocabulary
Listening	Retrieve (information, meaning of words, meaning of a text/conversation/discourse); Interpret (make inferences, generate conclusions, summarize main ideas, non-verbal language, communicative intentions); Reflect (opinion, purpose of different communicative situations)	
Speaking	Form (pronunciation, syntax); Content (amount and variety of vocabulary, coherence); Use (prosody, pragmatics)	
Vocabulary	Acquire new words (by context, through images and research); Recognize (relationships between words, elements that make up a word, linguistic varieties, everyday expressions from the area or from originary languages, word-formation, semantic categories)	
METALINGUISTIC COMPETENCY		Phonological awareness
Phonological awareness	Distinguish (onset and rhyme, words, syllables, phonemes); Blend (words, syllables, phonemes); Generate words from (rhymes, syllables, phonemes); Segment (in syllables, in phonemes, spelling)	

Figure 11. Content descriptors for mathematics (UNESCO-IBE and UIS, 2017)

MATHEMATICAL PROFICIENCY	Problem solving, reasoning, communicating and presenting arguments
Problem solving	Understand, plan, do and check mathematics tasks
Reasoning	Accurately represent problem elements, concepts and procedures; justify solution approaches
Argument and communication	use mathematical language, connect mathematical ideas to everyday life, and interpret mathematical statements
NUMBER KNOWLEDGE	Competency with pre-number ideas (e.g., counting), number ideas (e.g., cardinality) and number systems (e.g., fractions)
Pre-number concepts	Number sense (mainly through counting), working with pre-number operations (e.g., counting objects)
Number, numeration and number systems	Natural numbers, fractions, decimals, integers, sets, exponentiation and alternative number systems (e.g., base 5)
MEASUREMENT	Using non-standard (e.g., play money) and standard (e.g., mass) units to measure various things
Non-standard units	Measuring quantities in the local environment with non-standard units (e.g., pencil lengths, water glass volume)
Standard units	Measuring quantities in the local environment with standard units(e.g., cm, lbs, L)
STATISTICS AND PROBABILITY	Data management (e.g., organizing and representing data), chance and probability experiments (e.g., coin toss trials)
Data management	Collecting, representing and interpreting data (e.g., plan, collect and analyse students' height data)
Chance and probability	Simple experiments to investigate chance and probability (e.g., coin toss trials)
GEOMETRY	2-D figures and 3-D shapes (e.g., lines and angles, cubes), completing transformations (e.g., rotations, reflections), working in the Cartesian plane (e.g., plotting)
Shapes and objects	Constructions and symbolic argument (e.g., construct a circle, prove sum of angles triangle theorem)
Position and direction	Transforming geometric figures (e.g., translation, rotation)
Properties of space	Cartesian plane (e.g., plot coordinate pairs, transform figures in the plane)
ALGEBRA	Non-numerical patterns (e.g., patterns found in the environment), numerical patterns (e.g., sequences), functions (linear, non-linear) and variation (e.g., ratio, proportion)
Non-numerical patterns	Patterns in the local environment and culture (e.g., mosaics, Mayan calendar)
Numerical patterns	Numerical relations (e.g., sequences, algebraic expressions)
Functions	Linear and non-linear functions (e.g., properties of straight lines, parabolas, simultaneous equations)
Variation	Ratio, percentage and proportion