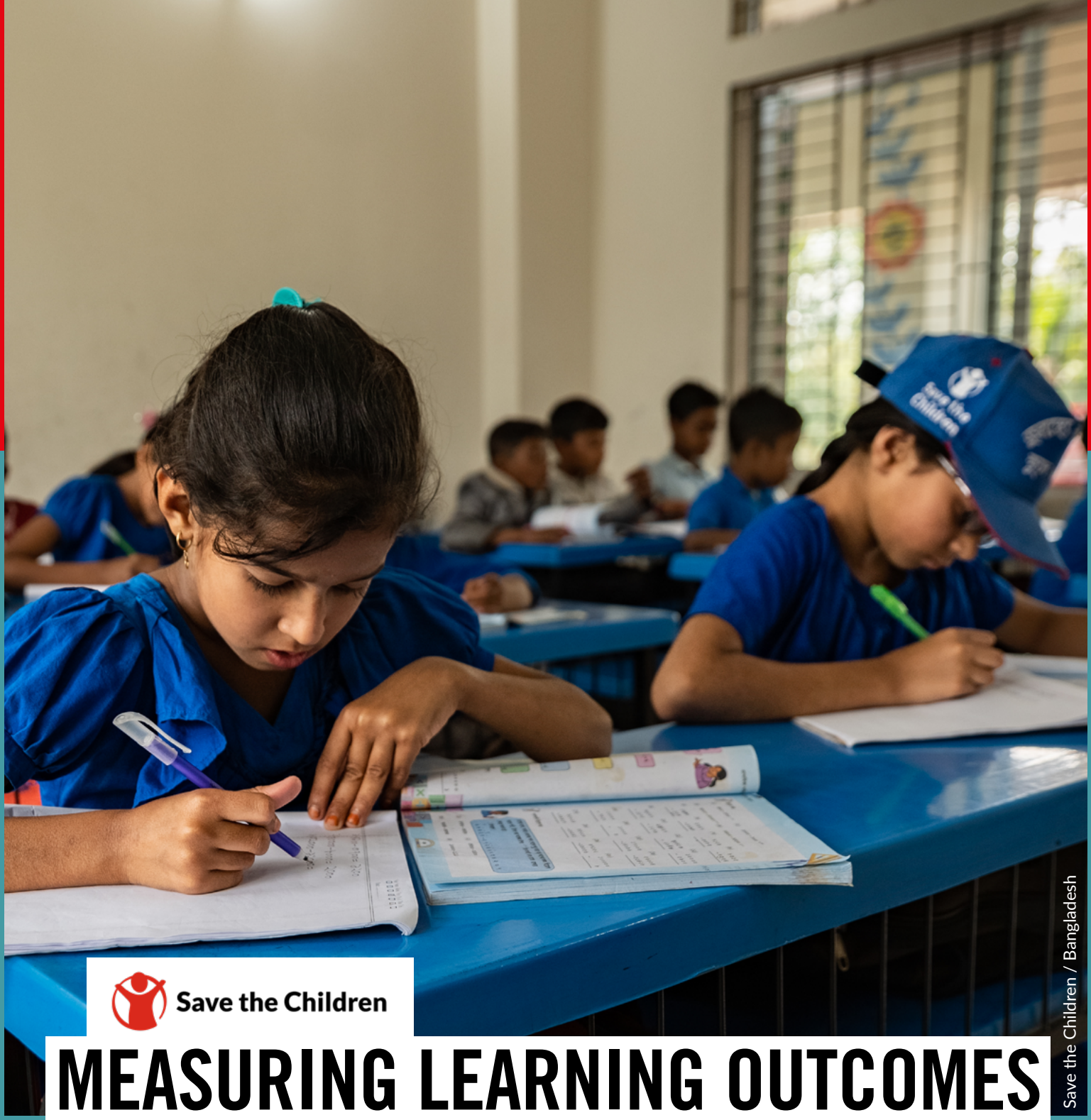


Save the Children / Bangladesh

Save the Children

MEASURING LEARNING OUTCOMES GUIDANCE

2024

PURPOSE

This guidance is written for education and MEAL technical experts working in both development and humanitarian contexts. The purpose of this guidance is to provide you with information on why we measure learning; how to choose the right assessment for programme evaluation, including whether or not you should adapt an existing assessment or create a new assessment; tips for planning and conducting assessments; and considerations for making the best use of data.

INTRODUCTION

Learning assessments are important for education programming **in all contexts**, **including emergency contexts**. Learning assessments in this guidance refers to data collection tools used—through direct observation, enumerator administration including teachers or caregivers, or self-reporting—to measure children's and adolescents' knowledge, attitudes, or skills related to literacy, numeracy, or social and emotional learning (SEL).[i]

Findings from learning assessments can support implementers, donors, and partners to make adjustments to existing programming and inform future programming. In crisis contexts, it may be challenging to implement feasible, rigorous, and contextually relevant assessments. Further, assessment systems often fail to capture or track the evolving needs of mobile populations. Despite these challenges, conducting learning assessments is critical to ensuring education programming is addressing the learning needs of children and adolescents, no matter who they are, where they are, or what is happening around them.

ACKNOWLEDGEMENTS

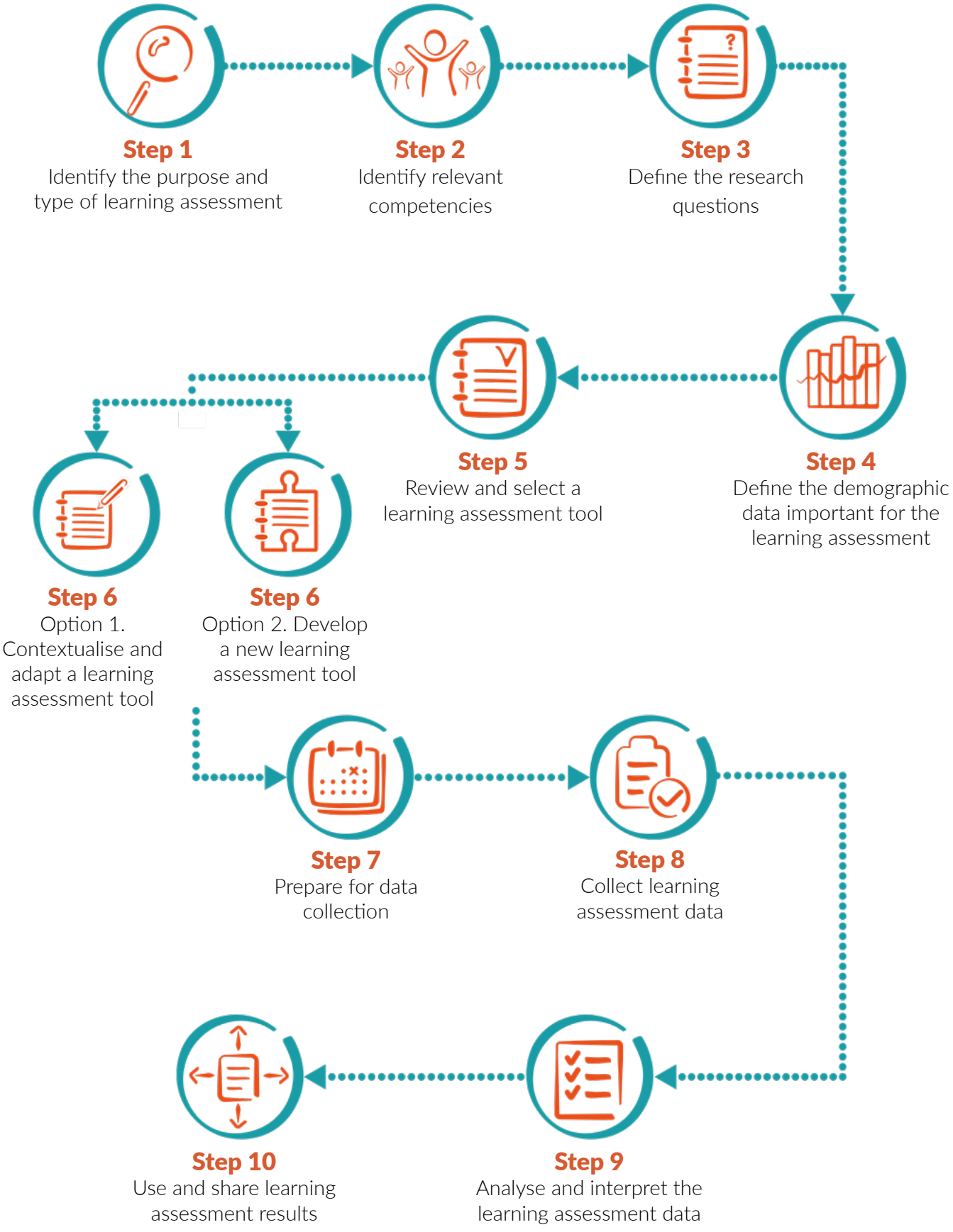
This guidance was originally written by Allyson Krupar, Kathryn Cooper, Simon Blower (Save the Children), Aimee Reeves, Emily Knowles-Crane, Carol da Silva, Briona Graham-Clayton, and Melanie Phillips (School-to-School International).

Following, the guidance was revised for publication by Stephen Richardson, Yaëlle Stempfelet, and Kathleen Denny (Wonder Lab) with support from Save the Children staff including Marian Hodgkin, Allyson Krupar, Kirsten Mucyo, Benjamin Hill, and Fergal Turner. The guidance was graphically designed by Edward Hodgkin.

i Wellbeing is central to Save the Children's definition of quality education. This guidance only covers the measurement of learning outcomes, specifically, literacy, numeracy, and SEL. SEL supports both learning and wellbeing. Save the Children is currently working on guidance to support measuring children's wellbeing outcomes from a broader perspective to complement this guidance on learning assessments

OVERVIEW OF THIS GUIDANCE

This guidance has 10 practical steps for how to plan, conduct, and learn from the results of a learning assessment, each with different considerations and resources.



OTHER TYPES OF EDUCATION-RELATED ASSESSMENTS NOT INCLUDED IN THIS GUIDANCE

This guidance focuses specifically on the role which learning assessments play in evaluating education programmes. There are a variety of other important purposes for measuring learning outcomes that this guidance does not cover including:

- **Situation analysis:** Can be used as a needs assessment at the population level to tailor initial programme design; cannot be used to generalise impacts of a programme, which may require a baseline and an endline.
- **Screening and placement:** Can be used to place children and adolescents in different programmes, grades, or for specific services; are targeted at the individual level and are generally not used to generalise impacts of a programme.
- **Tracking:** Formative classroom assessments help teachers focus on specific needs in the classroom; are targeted at the individual level and cannot be used to generalise impacts of a programme.
- **Monitoring:** Can be used for adaptive management and are cross-sectional; targeted at population level but are generally not used to generalise impacts on an entire population.
- **Longitudinal study:** Measures change over time longitudinally; targeted at population level and can be used to generalise impacts on an entire population.[1]

Save the Children / Sierra Leone

STEP 1. IDENTIFY THE PURPOSE AND TYPE OF LEARNING ASSESSMENT

The first step in planning a learning assessment for programme evaluation is to clearly define **why** you are conducting it. This helps you choose an appropriate assessment tool, design your assessment, including the sample of learners you will assess, and how often you will assess them. The three main purposes for learning assessments covered in this guidance are learning, accountability, and equity. You can select one or a combination of these purposes. Think carefully about which of these purposes are most important to the desired outcomes, and then as you read the other steps in this guidance note, reflect on how they are related to the purpose you identified.

LEARNING

Learning assessments measure learning outcomes to understand what learners know and how their literacy, numeracy, and SEL skills develop over time. You can disaggregate by locally-appropriate demographics, such as geographically, by school, and/or gender to understand the relationships between programme activities and learning outcomes and then compare results. You can also identify if there is a relationship between literacy or numeracy learning outcomes and SEL outcomes.

If you are conducting an assessment for learning, you will likely use assessment findings to adjust programme activities and other inputs, including curricula, to better meet the needs of learners participating in the programme.

ACCOUNTABILITY

Learning assessments for accountability provide feedback to project stakeholders such as learners, caregivers, teachers, communities, education authorities including governments, and donors on what children and adolescents learnt and how to better support their learning at school and at home.

If you are conducting an assessment for accountability, your focus will likely be on how well the project meets its performance goals, typically defined by a project's performance indicator targets. This information supports donors to evaluate the effectiveness of education programmes they invest in, how well partners performed, and report back to project stakeholders, including learners and caregivers.

Identifying relevant indicators

Programme evaluations require reporting on specific project indicators. These are in the project's logframe or results framework. The purpose of the indicator should match those of the learning assessment. Learning indicators may include the following types of measures:

- Mean assessment score overall or on specific subjects or domains within an assessment.
- Percentage of learners meeting a locally defined learning benchmark.
- Percentage of learners with improved learning in one or more subjects - this requires you to measure both a baseline and an endline.

In some instances, donors such as the United States Agency for International Development (USAID) and the United Kingdom's Foreign, Commonwealth, & Development Office (FCDO) may require you to use a specific indicator and affiliated learning assessment for the programme evaluation. See Annex 1: Illustrative Indicators.

EQUITY

Learning assessments for equity assess if and to what extent any inequities exist between different learner groups (e.g. gender, displacement status, nationality, ethnicity, disability, home language, etc.) If inequities in learning outcomes do exist, assessment findings can provide data that can help you identify barriers faced by different learner groups.

If you are conducting an assessment for equity, you can work to address any barriers that exist for learner groups, which you can use to inform future programme design. You can also share the findings with other education stakeholders to ensure they have the information needed to support all learners' needs.

Accountability, learning, and equity are important purposes for measuring learning in programme evaluations. Additionally, a learning assessment may serve multiple purposes. For example, the baseline assessment for an evaluation could also be used as a needs assessment and/or placement test if planned appropriately.

There are other programme evaluation purposes which are not covered in this guidance such as value for money, etc. (see the text box below and the other types of education-related assessments previously mentioned). You can use many of the considerations from this guidance to plan and conduct these other types of assessments, however, this document does not provide the full range of information that you may need.

Value for Money

Cost analyses can provide project stakeholders with information about a project's cost-efficiency, cost-effectiveness, or cost-benefits compared to other projects or specific components of the project. This is important when thinking about learning assessments, because we need to know what interventions are the best value for money to achieve learning outcomes. This may vary by context and equity, but in an ever-challenging funding climate, we need to ensure we are making the best use of resources to optimise learning outcomes. This contributes to global research and learning towards evidence based solutions for impact, like the World Bank's 'Smart Buys'. If you are considering conducting a programme evaluation that also includes understanding Value for Money (VfM), you will include project cost data in your analysis.USAID provides comprehensive guidance on cost measurement for education projects. FCDO has general guidance on evaluating VfM, and its Girls' Education Challenge has published more specific guidance on evaluating VfM for education.

STEP 2. IDENTIFY RELEVANT COMPETENCIES

Once you identify the purpose of the learning assessment, you will need to identify relevant competencies. Competencies include knowledge, attitudes, skills, and behaviours (see Table 1). A domain refers to a specific area or subject matter. The competency should align with the project purpose, activities, theory of change, and indicators, if applicable.

Table 1: Competencies and definitions [2 3 4]

Competency types	Competency definition
Knowledge	The ability to recall facts, information, and concepts.
Attitudes	The implicit or explicit beliefs that people hold and which influence how they act.
Skills	The ability to do things such as adding numbers, reading with fluency, cooperating with others to solve conflicts.
Behaviours	The way people act and conduct themselves.

For education programmes, common domains include literacy, numeracy, and SEL:[ii]

- **Literacy** competencies include, but are not limited to, listening comprehension, oral language, phonological awareness, alphabet knowledge, print awareness, reading fluency, and reading comprehension.[5] In multilingual settings, important language competencies to assess might include receptive and expressive vocabulary, morphology, and syntax.

- **Numeracy** competencies include number identification, number discrimination, number pattern identification, addition and subtraction, multiplication and division, fractions and decimals, geometry, and measurement, among others.[6]

- **Social Emotional Learning (SEL)** may be referred to as life skills, soft skills, 21st century skills, non-cognitive skills, and transferable skills depending on the context.[7 8] Each term uses a slightly different theoretical framework driven by the needs of a specific field. However, all these terms refer to social and emotional competencies that develop in a child's life. SEL competencies depend on programme design and cultural considerations. See the Inter-Agency Network for Education in Emergencies (INEE) and Harvard Graduate School of Education's Ecological Approaches to Social Emotional Learning (EASEL) Laboratory's PSS-SEL Toolbox as well as Harvard University's Explore SEL for valuable resources on SEL domains and tools. See also Annex 2: SEL Competencies.

ii In acute humanitarian contexts, SC is also exploring how to measure lifesaving learning (i.e. knowledge and skills that keep children safe in the face of specific hazards such as earthquakes, land-mines, or contagious diseases), with tools and approaches still at early stages of conceptualisation.

STEP 3. DEFINE THE RESEARCH QUESTIONS

TIP

Having only one or a limited number of research questions for the programme evaluation will make the analysis more straightforward. Programme evaluations that attempt to answer too many questions may be spread too thin and therefore not conclusive or of substance, and are expensive to conduct. If needed, you can work with a group of stakeholders to select and prioritise the research questions.

After identifying the relevant competencies for the learning assessment, you can identify the related research question(s) the programme evaluation intends to answer. The research questions chosen will inform the design of the assessment, including what tools you use, the demographic data you collect, and the sample for your study. This can have significant cost and logistical implications. For example, questions which look at programme impact will need to have a control group, and those which look to compare the effectiveness of different elements of a programme may require you to implement different versions of the programme for different populations. This may mean that in some cases you are limited in what research questions are feasible for you to answer.

Defining the research questions helps you to:

- Choose the right assessment for your needs,
- Collaborate with stakeholders to identify their interests,
- Avoid confirmation bias, ensuring that you seek results to unbiased questions rather than just confirming what you expect, and
- Learn about the project's strengths and challenges, allowing you to make necessary adjustments.

With that said, example research questions could include the following:[iii]

LEARNING

- Which parts of the project or intervention contribute the most to children's learning? To what extent?
- Did rapid response interventions contribute to improved foundational skills for children?
- How do learners' rate of acquisition of SEL skills compare to that of children of the same age not experiencing an emergency or crisis?

ACCOUNTABILITY

- Does this programme meet its intended goal of improving reading outcomes?
- Did our teacher training significantly improve children's SEL skills?

EQUITY

- Are learning outcomes different between children attending temporary learning spaces and host government schools?
- Are girls and boys benefiting equally from the new approach?
- Who are the highest and lowest-achieving learners year on year?
- What adaptations or accommodations have the greatest impact on learning for children with different forms of disabilities?

iii Save the Children staff can find the Global Research Evidence and Learning Agenda here, including questions relevant to learning assessment.

STEP 4. DEFINE THE DEMOGRAPHIC DATA IMPORTANT FOR THE LEARNING ASSESSMENT

TIP

When disaggregating for disabilities, use validated questionnaires or screening tools such as those published by the Washington Group on Disability Statistics. You should only disaggregate by disability if the learning assessment tool(s) administered to children with disabilities can be equivalent with adaptations for children without disabilities. If the assessment includes modifications, it is not appropriate to compare results across learners with different disabilities.

Capturing demographic data is crucial for analysing equity in outcomes across different contexts and learner subgroups. This data allows you to disaggregate results by various factors, some of which donors may require. When defining the demographic data it is important to review the research questions, institutional learning, past research, and donor requirements.

Some demographic data that you may need to capture as part of the learning assessment include:

- Gender
- Socio-economic status (SES)
- Caregiver information, including educational background
- Home language
- Age(s)
- Grade(s)
- Disability status
- Displacement status

Demographic data should include everything you need to answer your research questions. In turn, it will also shape how you approach sampling, as you may need to over sample certain groups, such as children with disabilities, to allow for a large enough sample to conduct meaningful analyses.

STEP 5. REVIEW AND SELECT A LEARNING ASSESSMENT TOOL

When selecting a learning assessment tool, first review the available options and consider those you can adapt or contextualise. This can include assessments previously used in the context or by another organisation.

When reviewing learning assessment tool options, ensure the tool aligns with the curricula used by the programme. If the curriculum aligns with national learning frameworks, the assessment should also align with national government assessments if they exist and are aligned to the national learning framework. However, in crisis and emergency contexts, or when services are provided by non-state entities, curricula may not always match national learning frameworks. To support your review and selection of a learning assessment, Table 2 includes a list of common validated assessments, some of which Save the Children developed.

Table 2: Common Validated Assessments

Tool	Description
Tools Developed by Save the Children	
International Development and Early Learning Assessment (IDELA)	**Developed by:** Save the Children **Outcomes measured:** Motor Development, Emergent Literacy, Emergent Numeracy, SEL Development **Description:** IDELA provides a holistic picture of children's development and learning in pre-primary years. It is a direct assessment for children aged 3.5 to 6 years. It can be adapted to a variety of income, development, and cultural contexts. It can be used for a variety of evaluation and research purposes, including programme evaluation.
Holistic Assessment of Learning and Development Outcomes (HALDO)	**Developed by:** Save the Children **Outcomes measured:** Literacy, Numeracy, SEL **Description:** HALDO is a learner assessment that measures literacy, numeracy, and SEL of children aged four to twelve years affected by crisis and emergency. It was initially piloted in 2018. It has an adaptive structure to account for a wide age range. HALDO can be used for programme evaluation.
International Social and Emotional Learning Assessment (ISELA)	**Developed by:** Save the Children **Outcomes measured:** SEL **Description:** ISELA is a scenario- and performance-based measure designed to assess the development of self-concept, stress management, perseverance, empathy, relationship management, and conflict resolution in children ages six to twelve years. ISELA can be used for programme evaluation.

Tool	Description
Tools Developed by Save the Children	
Holistic Assessment for Learning (HAL)	**Developed by:** Save the Children, UNICEF, New York University Global TIES for Children **Outcomes measured:** Literacy, Numeracy, SEL **Description:** HAL is a tool for measuring early grade learners' literacy, numeracy, and SEL. HAL is based on the Syrian curriculum for grades 2 and 3. The objective of HAL is to provide teachers with a rigorous formative assessment tool that they can use to monitor the progress of learners in their classrooms.
Literacy Boost Reading Assessment (LBRA)	**Developed by:** Save the Children **Outcomes measured:** Literacy **Description:** The LBRA is the first step in the Literacy Boost model. Assessments help projects, schools, and partners identify strengths and gaps in learners' literacy skills and monitor progress. Additionally, these assessments can be used to compare the progress of Literacy Boost learners to those who are not yet receiving Literacy Boost support. It can also be used as a summative assessment tool for programme evaluation.
Numeracy Boost Assessment (NBA)	**Developed by:** Save the Children **Outcomes measured:** Numeracy **Description:** The NBA helps projects, schools, and partners to identify strengths and gaps in learners' numeracy skills and monitor their progress. The NBA can also be used as a summative assessment tool for programme evaluation.
Remote Assessment of Learning (ReAL)	**Developed by:** Save the Children **Outcomes measured:** Literacy, Numeracy, SEL **Description:** ReAL assesses children's learning remotely, focusing on literacy, numeracy, and SEL. It can be administered by smartphone or regular mobile and there are three different adaptations (High Access, Low Access, Caregiver) dependent on what materials (stimuli) and types of phones are available.

Tool	Description
Tools developed by other entities	
Caregiver Reported Early Development Instruments (CREDI)	**Developed by:** Harvard University's Graduate School of Education **Outcomes measured:** SEL, Receptive and Expressive Language **Description:** CREDI is a set of population-level measures of early childhood development for children from birth to age three.
Measuring Early Learning Quality and Outcomes Child Development and Early Learning (MELQO-Model)	**Developed by:** MELQO core team **Outcomes measured:** Executive Function, SEL, Early Mathematics, Early Literacy **Description:** MELQO-Model measures children's learning and development through two tools—a direct assessment and a teacher or caregiver survey—designed to assess basic domains of children's development at the start of school.
Social-Emotional Response and Information Scenarios (SERAIS)	**Developed by:** International Rescue Committee, New York University Global TIES for Children **Outcomes measured:** SEL **Description:** SERAIS is a scenario-based learner assessment designed to measure SEL skills for elementary school-aged children. It has been used for programme evaluation purposes.
Multiple Indicator Cluster Survey: Foundational Learning Skills Module (MICS-FLS)	**Developed by:** UNICEF **Outcomes measured:** Literacy, Numeracy **Description:** MICS-FLS captures basic literacy and numeracy skills in grades 2 and 3, targeting children aged seven to fourteen years to monitor learning and quality of education. The module was developed for use in household surveys.
Annual Status of Education Report Assessment (ASER)	**Developed by:** Pratham and ASER Centre **Outcomes measured:** Literacy, Numeracy **Description:** The ASER is a literacy and numeracy assessment developed by the Indian organisation Pratham to empower community members in monitoring children's learning outcomes. It is customisable to different languages and contexts. The ASER assesses foundational levels of reading and numeracy. The ASER includes eight tasks—four literacy and four numeracy tasks—and can be used for various evaluation and research purposes, including programme evaluation.

Tool	Description
Tools developed by other entities	
Early Grade Reading Assessment (EGRA)	**Developed by:** RTI International **Outcomes measured:** Foundational Literacy **Description:** The EGRA is an orally administered instrument to report on learners' foundational literacy levels. It was first piloted in 2007 and has been used in over 70 countries. It generally includes five or six subtasks and can be customised to different languages and contexts. EGRA can be used for various evaluation and research purposes, including programme evaluation.
Early Grade Mathematics Assessment (EGMA)	**Developed by:** RTI International **Outcomes measured:** Foundational Numeracy **Description:** The EGMA is a simple, orally administered instrument to report on learners' foundational numeracy levels. It generally includes six subtasks and can be customised to different contexts. EGMA can be used for various evaluation and research purposes, including programme evaluation.
Secondary Grade Reading Assessment (SeGRA) and Secondary Grade Mathematics Assessment (SeGMA)	**Developed by:** FCDO's Girls' Education Challenge (GEC) **Outcomes measured:** Literacy, Numeracy **Description:** The GEC's SeGRA and SeGMA are custom-built assessment frameworks that bridge early grade testing and grade-appropriate curriculum-referenced testing. As SeGRA and SeGMA are designed as frameworks for developing curriculum referenced tools, the document linked here provides guidance on developing context specific SeGRA and SeGMA tools, rather than tools themselves. SeGRA and SeGMA subtasks are designed to test reading and mathematics skills progressively, approaching fully functional levels of literacy and numeracy in the final subtasks. SeGRA and SeGMA are paper-based, written assessments and can be used for programme evaluation.
YouthPower Positive Youth Development (PYD)	**Developed by:** YouthPower Learning **Outcomes measured:** SEL **Description:** PYD's measurement toolkit measures four domains promoted through the PYD framework: assets, agency, contribution, and enabling environment. It is used for programme evaluation.
Rugged Resilience Measure (RRM)	**Developed by:** Resilience Research Centre **Outcomes measured:** SEL **Description:** RRM is a self-report measure of internal or psychological resilience and measures capacities in essential skills and strengths that have been found to be protective against many forms of stress and adversity. It has been used for individuals ages 16 to 65.

ASSESSMENTS BY COMPETENCIES

When selecting an assessment tool, ensure it aligns with the competencies that need to be measured to answer the evaluation questions and supports programming. Ensure alignment between the assessment and what needs to be measured to understand project goals, objectives, and outcomes. For example, if the programme intends to improve primary school aged children's literacy and SEL, then you may need to combine a literacy assessment, like EGRA, with an SEL-focused assessment, such as ISELA, or choose an assessment tool that covers both like HALDO.

You should also consider the specific competencies that the assessment tool measures. See Figures 1 and 2 that illustrate the literacy and numeracy domains included in various learning assessment tools.

As mentioned, SEL competencies depend on the programme design and cultural dimensions. INEE's Assessment of Academic Learning Outcomes and Psychosocial Support and Social Emotional Learning Outcomes in Education in Emergencies provide valuable, in-depth summaries of these assessments and their characteristics. Additionally, INEE's Measurement Library delivers a wide range of tools to assess learning and holistic development in crisis and emergency contexts.

Save the Children / Burkina Faso

FIGURE 1: ASSESSMENT COMPETENCIES—LITERACY (ADAPTED FROM INEE)[9]

Tools	Grades/Ages	Reading		Linguistic			Metalinguistic	Writing
		Decoding	Comprehension	Listening	Speaking	Vocabulary	Phonological awareness	
CREDI*	ECCD (0-3)			■	■	■		
IDELA*	Pre-primary (3-5)	■	■	■	■	■	■	■
HALDO*	Pre-primary (3-5) Lower Primary (6-9) Upper Primary (10-13)	■	■	■	■	■	■	■
ASER*	Lower Primary (6-9) Upper Primary (10-13) Secondary (14-18)	■	■					
EGRA*	Lower Primary (6-9)	■	■	■	■	■	■	■
HAL*	Lower Primary (6-9)	■		■		■		
LBRA	Lower Primary (6-9)	■	■	■	■	■	■	■
MELQO-MODEL	Lower Primary (6-9)			■	■	■	■	■
MICS-FLS	Lower Primary (6-9) Upper Primary (10-13)	■	■			■	■	
ReAL	Lower Primary (6-9) Upper Primary (10-13)	■	■	■		■		
SeGRA	Upper Primary (10-13) Secondary (14-18) Lifelong (18+)	■	■			■	■	

■ Indicates that the tool includes the literacy subdomain.

*This tool has been used in crisis or emergency contexts.

FIGURE 2: ASSESSMENT COMPETENCIES—NUMERACY (ADAPTED FROM INEE)[9]

Tools	Grades/Ages	Math Proficiency**	Number Knowledge	Measurement	Statistics and Probability	Geometry	Algebra
IDELA	Pre-primary (3-5)	■	■	■		■	
HALDO*	Pre-primary (3-5) Lower Primary (6-9) Upper Primary (10-13)	■	■	■		■	
MELQO-MODEL	Pre-primary (3-5)	■	■	■		■	
ASER*	Lower Primary (6-9) Upper Primary (10-13) Secondary (14-18)	■	■				
EGMA*	Lower Primary (6-9)	■	■	■		■	■
HAL*	Lower Primary (6-9)	■	■			■	
MICS-FLS	Lower Primary (6-9) Upper Primary (10-13)		■				■
NBA	Lower Primary (6-9)	■	■	■		■	■
ReAL	Lower Primary (6-9) Upper Primary (10-13)	■	■				
SeGMA	Upper Primary (10-13) Secondary (14-18) Lifelong (18+)	■	■				

■ Indicates that the tool includes the numeracy subdomain.

*This tool has been used in crisis or emergency contexts.

**(e.g., problemsolving, reasoning)

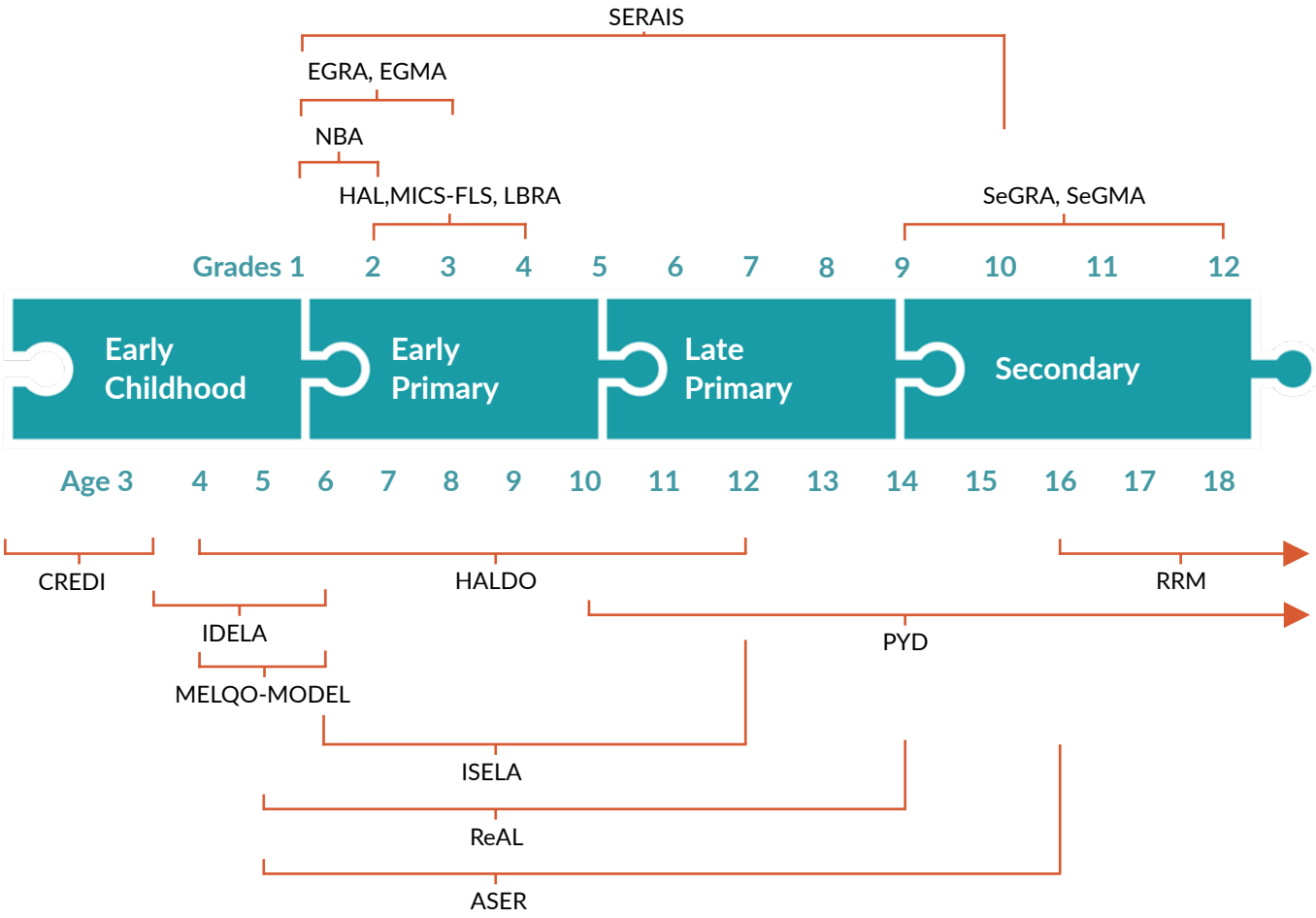
KEY CONSIDERATIONS WHEN SELECTING A LEARNING ASSESSMENT TOOL

Beyond ensuring the learning assessment tool is relevant to the purpose of the programme evaluation, there are a number of additional considerations when selecting a tool.

AGE AND GRADE LEVEL

Learning assessment tools are designed for specific age groups or grades, considering factors such as length, administration protocols, and developmental or curriculum-related content. When selecting an assessment tool, you need to consider the suitability of the tool for that age and grade range. Only tools developed and tested with learners in the same age range that you will be assessing should be used. See Figure 3 which illustrates different age and grade ranges for some common validated assessments. Note that age and grades do not always align. For example, if your project supports out-of-school children, the assessment may need to align with the grade level as opposed to the age level.

FIGURE 3: AGE AND GRADE RANGES OF ILLUSTRATIVE ASSESSMENTS (ADAPTED FROM INEE)[9]



LOCATION AND LANGUAGE

If the project's location and language match an existing version of a selected learning assessment tool, that assessment has likely already been adapted for local, child-friendly language. However, it should still be reviewed to make sure the language matches what each question intends to measure. When evaluating a specific **curriculum**, select tools that can be contextualised

to incorporate national learning frameworks, curricula, and contextually appropriate references, terminology, and images. When selecting a tool to measure SEL, it is especially important to consider location and language. SEL concepts are informed by social and cultural values and norms, and are highly context specific.

CHILDREN WITH DISABILITIES

If your project includes children with disabilities, first identify if the tool has already been adapted to assess children with disabilities in that context. If the selected learning assessment tool has not been used with children with disabilities, you will need to adapt the tool with the support of disability experts, organisations of people with disabilities (OPDs), and inclusive education teachers to appropriately accommodate all children.

TYPE OF ADMINISTRATION

While the type of administration—whether a direct assessment with a learner, self-report, or caregiver/adult report—may not determine which tool you select, it will significantly affect the logistics and the budget. For example, if the tool involves a caregiver reporting on a child, you will need to either bring caregivers to a central location or visit households for the assessment. Both options require different logistics than a direct assessment with a child, where enumerators can gather information at a school or an education centre.

RELIABILITY AND VALIDITY

If an assessment tool does not have evidence of reliability or validity with the population the project is supporting, you cannot be certain that the results from the learning assessment accurately represents what the learners know.

Reliability measures whether an instrument yields the same results across multiple administrations (test-retest reliability) or compares whether enumerators score learners the same way (inter-rater reliability).**Validity** refers to the extent an instrument appropriately measures the constructs it intends to measure. There are a number of different measures of reliability and validity - for more information, please consult this INEE Measurement Guidance document.

USE IN CRISIS AND EMERGENCY CONTEXTS

When reviewing a learning assessment tool, it is crucial to determine whether or not it has been used and validated for a crisis or emergency context. It is also important to consider in which phase the tool was used (e.g. acute, protracted). When selecting a tool for these contexts, you should look for tools that are relatively easy to use and contextualise, require less technical capacity to deliver, and can potentially be done remotely, etc.

If an assessment has not been used in a crisis or emergency context, it still may be applicable to the programme evaluation. However, you must contextualise and adapt the assessment with local experts. In addition, you should also conduct validity and reliability testing with the learner population.

STEP 6. OPTION 1
CONTEXTUALISE AND ADAPT A LEARNING ASSESSMENT TOOL

When a learning assessment tool has recently been used in your project's context, language of instruction, and with a similar population, you may only need to conduct a light review of the tool. However, it is always advisable to perform a quality check of the assessment's instructions, protocols, and content, even if in-depth contextualisation and adaptation are not required.

When contextualising and adapting a tool, you should ensure the process is well-documented. Engaging with local, national, and international stakeholders, such as government officials, teachers, and local technical experts, is key to adapting assessment tools effectively. Technical experts may include those with expertise in education assessments, literacy including local languages, numeracy, SEL, etc.

3EA's Measure Guidance On How To Contextualise An Assessment

- Update language and review translations.
- Review semantic, idiomatic, and conceptual meanings.
- Align any content with the programme.
- Identify assessment accommodations needed for learners with disabilities, if relevant.
- Check for overall face validity through cognitive testing with local experts, including children.

Save the Children / Northeast Syria

STEP 6. OPTION 2
DEVELOP A NEW LEARNING ASSESSMENT TOOL

You should only consider developing a new learning assessment tool if no existing tools meet your needs. Creating a new assessment tool requires substantial time and financial resources, and is not recommended unless absolutely necessary. Before pursuing this option, you should first evaluate whether you can contextualise and adapt an existing assessment or a combination of tools for the programme evaluation.

Save the Children / Uganda

STEP 7. PREPARE FOR DATA COLLECTION

Once you select or develop a learning assessment tool in alignment with the purpose of the programme evaluation, there are other important actions that you should do prior to data collection: budgeting, developing a sampling strategy, implementation planning, training enumerators, and piloting the tool.

BUDGETING

Budget plays a crucial role in the feasibility of your learning assessment design, so you should address it early in the process. Sample size, including the number of schools and learners, as well as the tool selected significantly impact the budget.

Use resources like the EGRA Budget Calculator from RTI International and USAID to estimate costs. This Excel-based tool is adaptable for various learning assessments but may require adjustments based on the assessment tool and which tasks (e.g. design, data collection, analysis) are handled internally or externally.

Deciding whether or not to use an external service provider or consulting group:

When deciding between using an external service provider that specialises in education evaluation or conducting it internally, both of which have budget implications, consider these key factors:

- **Expertise:** Service providers can have specialised knowledge in education evaluation.
- **Objectivity:** Service providers may be less likely to be biassed in their assessments.
- **Confidentiality:** If the programme evaluation is sensitive, the service provider may provide more confidentiality as they are not directly affiliated with the organisation.
- **Flexibility:** Service providers can offer more adaptable evaluation options in terms of scope, design, and duration of the evaluation.

Conversely:

- **Time and resources:** Service providers can add time burdens to internal operations through procurement and logistics planning. Working with an external assessment group requires a work plan developed at the beginning of the project.
- **Cost:** Service providers charge for services and these costs need to be in programme budgets from the start for implementers to benefit from external service provider's expertise.

SAMPLING

When conducting learning assessments, it is important to choose a sampling strategy that balances practicality with the need for representative data. Representative sampling ensures that the sample accurately reflects the targeted population. The evaluation can achieve this through random sampling, where every individual or location has an equal chance of being selected, or through stratified random sampling, which involves dividing the population into specific categories (such as urban vs. rural schools), and then randomly selecting from each category. This approach helps in capturing diverse perspectives and conditions, making the findings more comprehensive and reliable.

Alternatively, the evaluation can use purposive sampling when time and resources are limited. This method involves intentionally selecting samples based on specific criteria or knowledge about the population, such as focusing on schools that have been significantly impacted by a crisis. While purposive sampling is less rigorous than representative sampling, and its findings are not generalisable to the entire population, it can still provide valuable insights, especially in emergency situations. This approach allows for a more efficient, although less rigorous, data collection process which ensures that critical information is gathered even under constraints.

See Step 4 for demographic considerations that might impact the sampling strategy, including the need for control groups and oversampling for specific groups. For more information, also consult J-PAL's resource for guidance on conducting power and sample size calculations. For more guidance on sampling, refer to SC's Sample Size Calculator Decision Tool.

Save the Children / Afghanistan

IMPLEMENTATION PLAN

Create an implementation plan that outlines the key details for administering the learning assessment, including the when, where, and who of administering it. This plan should specify key dates, required resources, and responsibilities for each step, serving as a reference throughout the assessment process. Make sure to plan and define data collection procedures, such as managing participants, creating a conducive assessment environment, and maintaining accurate records to ensure effective data collection.

ASSESSMENT TIMING

Take into account the academic year's start and end dates, summer breaks, holidays, and exam schedules to prevent conflicts with the learning assessment. Schedule assessments in the morning when possible, as young children tend to have better focus and concentration earlier in the school day.

PLANNING WITH THE MINISTRY OF EDUCATION

Whenever possible, meet with Ministry of Education personnel and school leadership well in advance to introduce the learning assessment and implementation partners. Ensure they understand the purpose, activities, and timeline of the assessment, and address any concerns before starting. Notify selected schools well in advance and provide important details about their participation, the purpose of the assessment, and logistical needs for data collection, such as providing a quiet room.

CONSIDERATIONS FOR CRISIS AND EMERGENCY SETTINGS

In crisis and emergency settings, your implementation plan should address how you will reach children in school, out of school, and displaced children. Be mindful that crises can cause migration or other movements, making it difficult to track learners over time or reach comparable groups. Identify potential risks to data collection and participants, and develop mitigation strategies suited to the specific context, including how to protect, secure, and destroy identifiable data. Also, make sure that all tools are conflict-sensitive. Finally, the timeline should be conscientious about not overburdening children, caregivers, teachers, and administrators who are already in challenging situations.

Electronic Data Collection

It is becoming standard practice to collect data electronically due to its increased accuracy and efficiency. Paper assessments come with increased opportunities for mistakes such as enumerators skipping questions, recording inconsistent responses, or miscalculating sums and timed information. Electronic data collection, however, automates processes, reducing assessor errors, and data entry mistakes. Additionally, electronic data collection allows you to identify errors immediately, enabling teams to quickly follow-up with the enumerator who made the error. Common platforms like CommCare and KoBo Toolbox are widely used for data collection, with KoBo Toolbox being free and requiring only an online account.

CHILD SAFEGUARDING AND ETHICS

Ensure the learning assessment adheres to child safeguarding and ethical standards. Review the assessment child safeguarding protocols to ensure they are in alignment with Save the Children's safeguarding policies. During the enumerator training, include sessions on child safeguarding (with specific focus on considerations for gender and children with disabilities) to cover child protection issues, responsibilities, and reporting procedures. Make sure all assessment tools and administration protocols adhere to ethical research practices.

Ensure that consent and assent procedures for children are clear, understandable, and appropriate. Participants should have the right to opt out at any time and be given adequate time to engage fully in the assessment. Encourage children, adolescents, and caregivers to ask questions or seek help as needed during the assessment. See existing Save the Children guidance on ethical research with children.

TRAINING ENUMERATORS

To improve data collection, recruit enumerators from the regions and communities where data will be gathered. Familiarity with local languages, dialects, and customs helps build rapport and answer questions participants have effectively.

Train enumerators on the content and purpose of the tools, the data collection platforms used, research ethics, child safeguarding, and their roles. Include practical exercises and hands-on practice with the tools. Conduct regular feedback sessions throughout the training to address any issues.

PILOTING

Piloting learning assessment tools helps identify challenges and allows you to make improvements before full data collection. You should always pilot tools in new contexts to ensure they are well adapted.

Train the pilot team on the tool and procedures before they test the tool on a small group of participants similar to the main study group. Ask enumerators to note any confusing, duplicative, or incomplete protocols and suggest improvements. They should provide feedback and recommendations on how to phrase questions, record responses, and simplify the tool's use.

During piloting, ensure support materials like verification forms are clear and accurate. Conduct inter-rater reliability measures to check for consistent scoring and to familiarise data collectors with working in pairs for 10% of the sample.

STEP 8. COLLECT LEARNING ASSESSMENT DATA

When conducting a learning assessment, it is crucial to actively monitor data to ensure its completeness and quality throughout the collection process including the use of inter-rater reliability measures, and address issues as they arise.

DATA MONITORING

During the assessment, you should monitor data by reviewing it for completeness and quality. Address any issues with data collection teams immediately to prevent future problems. Resolve any administration or assessment issues throughout the data collection period to maintain data quality.

ITEMS TO CHECK

You should verify that learners have unique IDs, assessment dates, school names, the correct number of assessments per grade or class, assessor names, and start and end timestamps of assessment records. Ensure enumerators follow time protocols by checking auto-stop times, and follow-up on any partial records.

INTER-RATER RELIABILITY

Inter-rater reliability measures reflect the reliability and consistency of data collected. As part of data collection, you should ensure teams collect inter-rater reliability data. Inter-rater reliability data is collected when two enumerators work in pairs to assess the same child, one speaking and marking responses while the other listens and marks responses. At least 10% of the sample should be assessed in pairs, generally the first child of each day. Enumerators can discuss their scoring, but cannot change their scoring after they assess the child together.

At the end of each day, data collection managers or analysts must review inter-rater reliability data and compare the scores to see if enumerators continue to make the same scoring decisions throughout the same assessment, or if there is disagreement. Any disagreement discovered during data collection should be discussed with enumerators.

STEP 9. ANALYSE AND INTERPRET THE LEARNING ASSESSMENT DATA

After you successfully collect the data, it is time to analyse and interpret it. An important first step is data cleaning, followed by analysis and interpretation that responds to the programme evaluation research questions.

DATA CLEANING

Clean the data by removing incomplete or unnecessary data before analysis. Document all instances of data removal and the reasons behind removal to ensure there is a clear path between the raw data that enumerators submit to the data you use for the analysis. If you intend to assess the same children at baseline and endline or at multiple points in time, you must have unique ID numbers associated with each child. These can be auto-generated or existing ID numbers from school registries, but they must be unique. Without unique ID numbers, you will not be able to track the same child over time.

DATA ANALYSIS AND INTERPRETATION

Data analysis highly depends on the learning assessment tools and associated methods. Some tools have procedures and/or protocols that are specific to that particular learning assessment to support the analysis process. Further, the analysis and interpretation will depend on the purpose of the programme evaluation and associated research questions. For example, if one of the purposes is equity, the analysis would assess the learning outcomes of the selected competency and then compare the results across different demographic data collected as part of the assessment (e.g. gender). In the implementation plan, ensure that the analyst responsible for the analysis and interpretation has adequate time to complete this process.

STEP 10. USE AND SHARE LEARNING ASSESSMENT RESULTS

How you use a learning assessment's results depends on the identified purpose (see Step 1), however, the primary use of results is to adjust and improve programming to better serve learners' needs and their communities.

Projects often share learning assessment results with donors and other external stakeholders, but rarely with learners, caregivers, teachers, or schools. However, it is essential to prioritise sharing results with all involved stakeholders, ensuring that data is anonymised and no identifiable information is shared, to protect learners and communities from potential harm. To improve usability, you should tailor results to each audience, offering guidance on interpreting results and actionable steps.

Consider the following six principles when sharing learning assessment results to ensure diverse stakeholders can understand and effectively use those results.

Six Principles for Sharing Learning Assessment Results

Once you have results on learning outcomes, you will want to present the results to a range of audiences, from programme teams, to cluster or government partners, to teachers, families, and learners themselves. While each of these audiences is different, there are some common principles which you can consider when presenting learning outcomes data.

1. Work from a strengths perspective.

While there is a global learning crisis, it is important to avoid using deficit-focused language when talking about learning outcomes, particularly at school and child level. Your learning data will show what learners can do, as well as what they cannot yet do. Focus on what learners are currently able to do, and how that is a starting point for what they can learn in the future. This keeps the focus on practical, positive actions.

2. Talk about learning in meaningful terms.

Data can often seem very abstract, particularly for audiences who do not regularly engage with data such as teachers or caregivers. Instead of presenting mean scores on assessments, use data which can be broken down into scores on specific sub-skills, and talk about what learners can do in practical terms. For example, instead of saying that learners scored a certain percent in reading, you can break it down into how many students have mastered different sub-domains of reading like letter recognition, word recognition, and so on.

Six Principles for Sharing Learning Assessment Results

3. Consider your audience and adapt your presentations.

People like to engage with information in a variety of ways. More technical audiences may want specific details on significance of results, but this may not be meaningful for others. As a general rule, focus on fewer data points, and try to present them both in writing and visually to suit a variety of audiences. It can also be useful to use "social maths" in which we talk about proportions of people (e.g. 7 out of 10 learners) instead of percentages (68.5 percent of learners).

4. Think before you compare.

Using learning data to rank and compare schools, districts, or countries may provide valuable insights for programme teams, however, for some other stakeholders, it may perpetuate negative sentiments and can distract from the real message of your presentation. When putting together information, consider why you are presenting comparisons, what research question it is answering, and whether there is a more neutral way to do it. For example, instead of comparing schools to each other, compare each school to the overall average, or even better, to a target benchmark.

5. Provide tools and space to link data with practice.

Engaging with data should be an active process in which what is learned from learning assessments is aligned with experience and evidence from other sources. Ideally, there should be space for discussion and validation of the assessment data with an opportunity to revise final reports based on stakeholder feedback. Where possible, compare learning outcomes with data on other outcomes, such as quality of learning environments.

6. Use a Universal Design approach.

Consider using a universal design approach to ensure comprehension across varying stakeholders and audiences, especially in terms of literacy levels and languages. Key strategies associated with universal design include:

- Sharing results in a way suited to each group's language and literacy levels,
- Following data protection protocols by anonymising participants, and
- Providing opportunities for stakeholders to ask questions and engage with the data.

When sharing results with children in particular, you should also ensure that the manner in which you present results is done in an engaging and child friendly way.

WHERE TO GO FOR MORE INFORMATION

For SC staff, for more information and support regarding planning learning assessments, contact: learningassessment@savechildren.org

For non-SC staff, consider reaching out to your local education coordination groups (Cluster or other), and/or the INEE Helpdesk.

Save the Children / Northeast Syria

REFERENCES

1. Krupar, A. (2023). *SEL and Wellbeing Outcomes: Tool Selection and Data Analysis* [PowerPoint Slides]. London: Save the Children.

2. Ferráns, S., & Lee, J. (2019). *Measure Guidance: Choosing and Contextualizing Assessment Measures in Educational Contexts*. New York: International Rescue Committee.

3. Markus, L., Thomas, H. C., & Allpress, K. (2005*). Confounded by competencies? An evaluation of the evolution and use of competency models*. New Zealand Journal of Psychology, 34(2), 117.

4. Zins, J. E., Weissberg, R. P., Wang, M. C., & Walgerg, H. J. (2004). *Building academic success on social and emotional learning: What does the research say?* New York: Teachers College Press.

5. RTI International. (2016). *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. Washington, D.C.: United States Agency for International Development.

6. Platas, L. M., Ketterlin-Gellar, L., Brombacher, A., & Sitabkhan, Y. (2014). *Early Grade Mathematics Assessment (EGMA) Toolkit*. Washington, D.C.: United States Agency for International Development.

7. Jones, S. M., & Doolittle, E. J. (2017). Social and Emotional Learning: Introducing the Issue. *Future of Children*, *27*(1), 3–12.

8. Save the Children. (2022). *The Quality Learning Framework, 2nd edition*.

9. Inter-agency Network for Education in Emergencies. (2020). *Mapping Exercise: Assessment of Academic Learning Outcomes*. New York: INEE.

ANNEX 1: ILLUSTRATIVE INDICATORS[iv]

Indicator name	Indicator level	Source
Percentage of children with an increased ability to use SEL (e.g., self-awareness, social awareness, responsible decision-making), critical thinking skills, creative thinking skills, stress management, problem-solving, conflict resolution, goal setting, collaboration or social skills, empathy, and emotional management via SEL or Life Skills classes	Outcome	Education Cannot Wait
Percentage of children with improved SEL (e.g., self-awareness, social awareness, responsible decision-making) outcomes compared with baseline	Outcome	Education Cannot Wait MHPSS indicator library
Percentage of children with an increased ability to display empathy towards others via focused psychosocial support	Outcome	Education Cannot Wait MHPSS indicator library
Percentage of children with an increased ability to display empathy towards others via SEL or Life Skills classes	Outcome	Education Cannot Wait MHPSS indicator library
Percentage of children with an increased ability to manage difficult emotions via focused psychosocial support	Outcome	Education Cannot Wait MHPSS indicator library
Percentage of children with an increased ability to manage stress via focused psychosocial support	Outcome	Education Cannot Wait MHPSS indicator library
ES.1–1 Percent of learners who attain a minimum grade-level proficiency in reading at the end of grade 2 with United States Government (USG) assistance	Outcome	USAID standard indicator
ES.1–2 Percent of learners attaining minimum grade-level proficiency in reading at the end of primary school (or grade 6, whichever comes sooner) with USG assistance	Outcome	USAID standard indicator
ES.1–47 Percent of learners with a disability targeted for USG assistance who attain a minimum grade-level proficiency in reading at the end of grade 2	Outcome	USAID standard indicator
ES.1–48 Percent of learners targeted for USG assistance with an increase of at least one proficiency level in reading at the end of grade 2	Outcome	USAID standard indicator
ES.1–54 Percent of individuals with improved reading skills following participation in USG-assisted programs	Outcome	USAID standard indicator
Number of crisis or emergency-affected boys, girls, and adolescents (3 to 18-year-olds) supported by Education Cannot Wait (a) in grades 2 or 3, or (b) at the end of primary, or (c) at the end of lower secondary education who achieve at least a minimum proficiency level in (i) reading and (ii) math	Outcome	Education Cannot Wait
EG.6–13: Percent of individuals with improved soft skills following participation in USG-assisted programs	Outcome	USAID standard indicator
Percentage of children aged 24 to 59 months who are developmentally on track in at least three of the following four domains: a) literacy, b) numeracy, c) physical, d) SEL	Outcome	INEE

iv Save the Children staff can find the Global Indicator List and guidance here, including additional SC-specific indicators and many of these donor-specific indicators.

ANNEX 2: SEL COMPETENCIES

Source	SEL competencies
CASEL Framework (Collaborative for Academic, Social, and Emotional Learning, 2024)	**Self-awareness:** The ability to understand one's own emotions, thoughts, and values and how they influence behaviour across contexts; includes capacities to recognise one's str engths and limitations with a well-grounded sense of confidence and purpose
	Self-management: The ability to manage one's emotions, thoughts, and behaviours effectively in different situations and to achieve goals and aspirations; includes capacities to delay gratification, manage stress, and feel motivation and agency to accomplish personal and collective goals
	Responsible decision-making: The ability to make caring and constructive choices about personal behaviour and social interactions across diverse situations; includes capacities to consider ethical standards and safety concerns and to evaluate the benefits and consequences of various actions for personal, social, and collective well-being
	Relationship skills: The ability to establish and maintain healthy and supportive relationships and to effectively navigate settings with diverse individuals and groups; includes capacities to communicate clearly, listen actively, cooperate, work collaboratively to problem solve and negotiate conflict constructively, navigate settings with differing social and cultural demands and opportunities, provide leadership, and seek or offer help when needed
	Social awareness: The ability to understand the perspectives of and empathise with others, including those from diverse backgrounds, cultures, and contexts; includes capacities to feel compassion for others, understand broader historical and social norms for behaviour in different settings, and recognise family, school, and community resources and supports
Save the Children Life Skills for Success (Moorcroft, Paruzzolo, & Press)	**Communication skills:** To actively listen and effectively express ideas and ambitions to different audiences and for different purposes
	Higher-order thinking skills: Ability to problem solve, think creatively, and make objective decisions
	Positive self-concept: Emotional, social, and cognitive awareness and positive valuation of oneself
	Self-control: Effectively regulate and manage behaviour and emotions, control impulses, and delay gratification to achieve personal goals
	Social skills: Ability to collaborate, resolve conflict, build respect and empathy for others, and behave in contextually appropriate ways; includes conflict resolution, prosocial behaviour, making and keeping friends, and identifying social norms
UNICEF Transferable Skills (UNICEF, 2019)	**Cognitive skills:** Related to "thinking" and the ability to focus, problem solve, make informed choices, and set plans and goals
	Social skills: Related to interactions with others, including the ability to communicate, collaborate, resolve conflicts, and negotiate
	Emotional skills: Related to understanding and regulating one's own emotions, including the ability to cope with stress, understand the emotions of others, and empathise with others
World Bank Step by Step Toolkit (World Bank, 2018)	**Self-awareness:** Knowing, understanding, and trusting ourselves; includes subskills of self-concept, self-efficacy, and emotional awareness
	Self-regulation: Governing our impulses and emotions; includes subskills of emotional regulation, delayed gratification, and frustration tolerance
	Social awareness: Understanding other people's feelings, needs, and concerns; includes subskills of perspective-taking, empathy, and prosocial behaviour
	Positive communication: Interacting with kindness and respect for ourselves and others; includes subskills of active listening, assertiveness, and conflict management
	Determination: Pursuing goals with resolve and purpose; includes subskills of achievement motivation, perseverance, and stress management
	Responsible decision-making: Making constructive and respectful choices; includes subskills of creative thinking, critical thinking, and responsibility

Source	SEL competencies
OECD Social and Emotional Skills Assessment Framework (Organisation for Economic Co-operation and Development, 2021)	**Open-mindedness:** Openness to experience; includes curiosity, tolerance, and creativity
	Task performance: Conscientiousness; includes responsibility, self-control, and persistence
	Engaging with others: Extraversion; includes sociability, assertiveness, and energy
	Collaboration: Agreeableness; includes empathy, trust, and cooperation
	Emotional regulation: Emotional stability; includes stress resistance, optimism, and emotional control
	Other: Includes achievement motivation and self-efficacy
Rwanda Basic Education Board Social-Emotional Learning Framework (Rwanda Basic Education Board, 2022)	**Thinking:** Includes cognitive processing and critical analysis
	Emotion: Includes emotional awareness and self-management
	Self-identity: Includes self-esteem and self-awareness
	Self-direction: Includes self-reliance and purpose
	Communication and understanding: Includes understanding, responding, and empathy
	Socialisation: Includes cooperation and relationship-building
	Citizenship: Includes conflict management and civic engagement

ANNEX 2 REFERENCES

Collaborative for Academic, Social, and Emotional Learning. (2024, January 22). *What is the CASEL Framework*. Retrieved from Fundamentals of SEL: https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/#schools

Moorcroft, S., Paruzzolo, S., & Press, S. (n.d.). *Save the Children Life Skills for Success*. London: Save the Children.

UNICEF. (2019). *Global Framework on Transferable Skills*. New York: UNICEF.

World Bank. (2018). *Step by Step - Toolkit Promoting Social and Emotional Learning (SEL) in Children and Teens*. Washington, D.C.: World Bank.

Organisation for Economic Co-operation and Development. (2021). *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills*. Paris: OECD Publishing.

Rwanda Basic Education Board. (2022). *Social-Emotional Learning Framework for Pre-Primary and Lower Primary Levels in Rwanda*. Kigali: REB.