



# ASSESSMENT-INFORMED INSTRUCTION:

# Systems Level

## Introduction

What is measured can be improved. For a long time, the world was focused on measuring access to education. As a result, we saw dramatic improvements in the number of children in classrooms. However, the measurement of learning outcomes has lagged. As of 2017, 100 countries were not systematically assessing student learning outcomes. Where we do have evidence, it is sobering: according to the United Nations, 617 million primary and lower-secondary school children (or 55% of all school children) are not minimally proficient in reading or mathematics. This problem is particularly acute in sub-Saharan Africa, where more than 87% and 84% of students were not proficient in reading or math in 2015, respectively.<sup>1</sup>

The World Bank has argued that to tackle this crisis, “it is necessary—though not enough—to measure learning.” **For assessment to inform instruction and improve learning outcomes, it must be followed by action.** Measurement allows governments to identify who is learning and who is not so they can target support to the schools and populations with the greatest need and ensure equity in learning support. It also enables them to identify what works to increase learning outcomes and how instruction might be improved.

Assessment-informed instruction has the potential to have significant impact on instructional quality and improved learning, but the evidence is somewhat limited. Studies have shown a modest relationship between high-quality assessments at different levels of the system and higher learning outcomes. Large scale learning assessments can be used to hold the system accountable. Countries with high-quality exit examinations have been shown to have higher learning outcomes than those that do not, and formative classroom assessments have been linked to improved learning outcomes.<sup>2</sup>

Yet only 17 out of 48 countries in sub-Saharan Africa have recently completed high-quality large-scale assessments.<sup>3</sup> Several countries have a variety of assessments in place, including national learning assessments, and some donors have sought to support governments in the development of system-level assessment structures, but few countries have been able to integrate system-level assessments into a broader foundational literacy and numeracy (FLN) learning program.

### BOX 1. Definitions

**Assessment in the system:** Measurement of teaching and learning outcomes across different levels of an education system

**System-level assessments:** Assessments that seek to evaluate the effectiveness of an education system

**Foundational literacy and numeracy:** The ability to read with meaning and do basic math calculations by grade 3

**Large-scale learning assessments:** National or cross-national standardized testing that provides a snapshot of learning achievement

**Content standards:** Expectations of what students at a given grade level should be taught

**Performance standards:** Descriptions of what students should be able to do

**High-stakes exams:** Exams that determine whether a student will advance to the next level

**Benchmark:** Specific assessment score that shows that a standard has been met

**Target:** A goal for the percentage of children who will reach the benchmark



## Guide Purpose and Structure

This guide describes how policy makers, donors, and implementing partners can design or refine assessments in the education system and connect those results to daily instruction. These activities will allow the system to obtain data to inform instruction, support program adaptation, ensure meaningful, non-punitive accountability, and link large-scale assessment results to what teachers teach every day. The purpose of assessment-informed instruction is to identify where learning gaps exist and target resources and improve the quality of instruction to address those gaps. This includes the gaps between learning outcomes and actual student achievement, as well as gaps between groups of children due to systemic inequities (such as gender gaps, gaps between ethnic groups, etc.).



This guide is organized as follows: First, it discusses how to analyze the current assessment context in your country. It then describes the various levels of assessment that exist in many systems, as well as the types of assessments that might be useful for these different levels. Next, it explores the cost considerations linked to these assessments. Finally, it provides a series of recommendations to increase the return on assessment investments in terms of improved learning outcomes.

### ANALYSIS OF THE CURRENT ASSESSMENT SYSTEM

Before embarking on a campaign to improve the use and sustainability of assessments, it is important to conduct a review of what assessments are currently being used in the context and how they are related to instruction. This review should include an analysis of the following:

**What assessments currently exist:** This includes looking at both “what should be” and “what is.” To identify “what should be,” start by conducting a desk review of laws and government policies regarding assessments and inspections, if they exist. Seek to find out at which levels assessments are mandated and how these existing assessments are linked to improving instruction. To identify “what is,” this review should include an examination of technical reports from the most recent assessments in the system, government inspection reports, results from the World Bank’s SABER data, and data from the country’s education management information system (EMIS). Follow the desk review with interviews with the Ministry of Education, inspectorate (if one exists), and any other government agencies involved in conducting assessments and inspections in order to identify how often the assessments are occurring on the ground, whether they are sample-based or census assessments, how they are funded, and who is conducting them.

**How assessments are used:** It is helpful to understand how assessment results are currently being disseminated and used, especially whether they are being used to identify where learning gaps exist in an effort to address those gaps. In this regard, you should examine several assessment levels:

- **International level** – e.g., for reporting to donors or on the Sustainable Development Goals
- **National level** – e.g., for informing policy and curricula, targeting resources, or advancing students
- **Subnational level** – e.g., for further targeting of resources and support, such as training, coaching, and inspection visits
- **School and classroom level** – e.g., for informing changes to pedagogy, prioritizing teaching time, or targeting differentiated instruction and acceleration programs

Gathering this information will require interviews with the government organizations responsible for disseminating and using results—to identify whom they distribute results to, how, and when. These information-collection efforts should not be just a paper survey; conduct interviews or focus group discussions with government officials and teachers. **Be sure to ask whether they have seen the results, what the results mean to them, how they used the results, and what might make the results easier to understand and use.**

Examine the alignment between the assessment, curriculum, and instruction (as described in more detail under point 3 of the section on Improving System-Level Assessment below).

**What informs teacher behavior:** It is also important to understand what currently drives teacher behavior and instructional choices, particularly how these elements intersect with the assessment system. You want to know



how teachers decide what to put in their daily lesson plans, what teaching and assessment methods they use, and how examinations affect their pedagogical choices. For example, what determines whether teachers are perceived to be successful? Do teachers feel pressure to “teach to the test,” and are those tests aligned with ensuring learners’ conceptual understanding of key FLN skills? How are teachers supported on key instructional and assessment techniques? This information can be collected from teachers through interviews, focus group discussions, or surveys.

## UNDERSTAND THE LEVELS OF ASSESSMENT

Countries generally have some assessment structure that exists at various levels of the education system, from the international to the regional, national, subnational, community, and school levels, as shown in Figure 1. To ensure a coherent assessment system, it is critical that assessments at these different levels be aligned with and build upon one another. We discuss the levels and types of assessments in further detail below. This guide focuses on assessments at the school level and above. The separate [Assessment-Informed Instruction: Classroom Level guide](#) focuses on the classroom level, providing guidance on how to develop low-stakes and simple measures of assessment that simplify the task of understanding learning levels without increasing the burden on teachers.

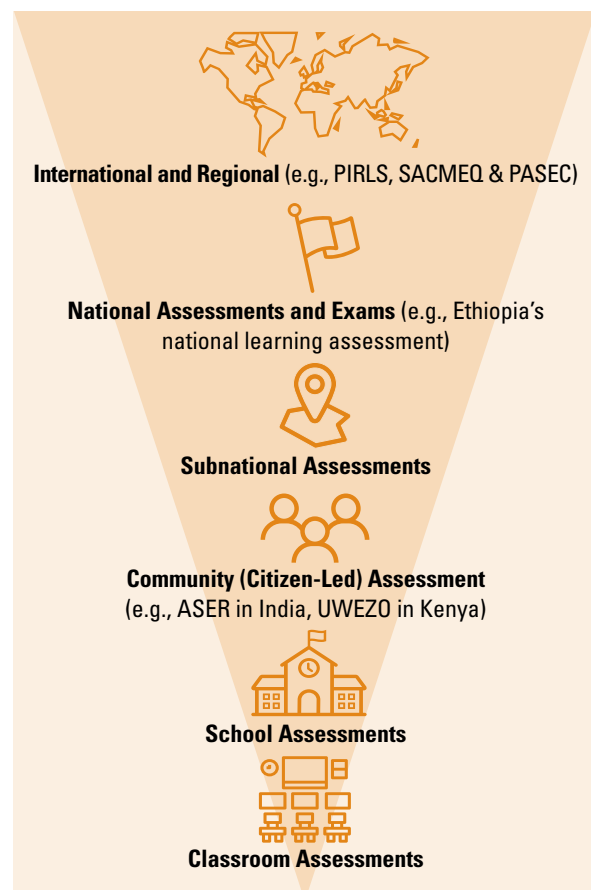
### OVERVIEW OF LARGE-SCALE LEARNING ASSESSMENTS

There are several types of large-scale learning assessments (LSLAs), which are used at the different levels of the assessment system (see Table 1). Consider which of these LSLAs might be appropriate in your country context and how they can work together to reduce costs while maximizing utilization for instructional decision-making.

### INTERNATIONAL AND REGIONAL ASSESSMENTS

*Countries should consider international and regional assessments since they provide opportunities to increase the comparability of outcomes with other countries, including utilizing learning-adjusted years of schooling (LAYS).* These assessments tend to be high-quality, reliable, transparent, and low-stakes paper-and-pencil-based exams that allow countries to compare their learning outcomes over time and with those of other countries. However, given the need to be relevant across country contexts, they come later in the education cycle (usually at grade 4 or above) and may not align well with a specific country’s curricula. They can also be substantially more costly than other assessments, and their results may take years. International and regional assessments do work well, however, at developing in-country assessment capacity and can highlight the need for curriculum reform or identify broad needs. For instance, in Uganda, the Ministry of Education set up a gender unit following SAQMEC results showing differential impacts by gender, and in Namibia, officials developed a policy to share good practices and resources between clusters of schools, after SAQMEC revealed deficiencies in this regard.<sup>4</sup> As with other types of assessments, **ensure that results from international assessments are shared with actors from various levels of the education system, with key recommendations for action.** Too often, teachers do not have opportunities to revise their pedagogy based on these assessment findings. FLN programs should review the findings from these assessments carefully to respond to the challenges identified in the assessments.

FIGURE 1. Levels of assessment




**TABLE 1. Types of large-scale learning assessments**

ASSESSMENT TYPE	OVERVIEW	EXAMPLES	GEOGRAPHICAL COVERAGE	TARGET POPULATION	SUBJECT(S)
International assessments	School- or sample-based, group-administered, standardized assessments meant to allow comparisons of student achievement across countries and over time <sup>5</sup>	PISA <sup>6</sup>	Worldwide, mostly OECD countries <sup>7</sup>	15-year-olds <sup>8</sup>	Reading, math, and science <sup>9</sup>
		PISA-D <sup>10</sup>	Low-and middle-income countries (LMICs) <sup>11</sup>	15-year-olds in grade 7+ and 14- to 16-year-olds who are out of school or enrolled in grade 6 or under <sup>12</sup>	Reading, math, and science <sup>13</sup>
		TIMSS <sup>14</sup>	Worldwide, mostly OECD countries <sup>15</sup>	Grades 4 and 8 <sup>16</sup>	Math and science <sup>17</sup>
		PIRLS <sup>18</sup>	Worldwide, mostly OECD countries <sup>19</sup>	Grade 4 <sup>20</sup>	Reading <sup>21</sup>
		LaNA <sup>22</sup>	LMIC countries <sup>23</sup>	Grade 4, 5, or 6 <sup>24</sup>	FLN <sup>25</sup>
Regional assessments	School- and sample-based, group-administered, standardized assessments involving a number of countries in a region <sup>26</sup>	PASEC <sup>27</sup>	French-speaking African countries <sup>28</sup>	Grades 2 and 6 <sup>29</sup>	Reading and math <sup>30</sup>
		LLECE <sup>31</sup>	Latin America <sup>32</sup>	Grades 3 and 6 <sup>33</sup>	Reading, writing, math, science, national factors that affect learning (cultural diversity, violence, etc.) <sup>34</sup>
		SACMEQ <sup>35</sup>	Southern and Eastern African countries <sup>36</sup>	Grade 6 <sup>37</sup>	Reading and math <sup>38</sup>
National exams	Group-administered high-stakes exams used to make certification and promotion decisions <sup>39</sup>	Many <sup>40</sup>	National or subnational <sup>41</sup>	End of primary school; end of lower secondary; university entrance exams <sup>42,43</sup>	Varies by country, but most include reading and math at a minimum <sup>44</sup>
National and subnational assessments	Vary, but generally school- and sample-based, group-administered assessments intended to assess whole-system quality <sup>45,46</sup>	Nepal's NARN; Ethiopia's NLA <sup>47</sup>	National or subnational, respectively <sup>48</sup>	Generally grade 4 and above <sup>49</sup>	Varies by country, but most include reading and math <sup>50</sup>
Early-grade reading and early-grade math assessments	School- and sample-based, individually administered assessments measuring FLN intended to act as a system diagnostic <sup>51</sup>	EGRA <sup>52</sup>	LMICs <sup>53</sup>	Grades 1–3 (and sometimes 4) <sup>54</sup>	Reading <sup>55</sup>
		EGMA <sup>56</sup>	LMICs <sup>57</sup>		Math <sup>58</sup>
Community-based and household assessments	Household- and sample-based, individually, volunteer-administered assessments that measure FLN and are intended to inform community perception <sup>59</sup>	ASER <sup>60</sup>	LMICs <sup>61</sup>	Children aged 5–16 <sup>62</sup>	Reading and math <sup>63</sup>
		ICAN <sup>64</sup>	13 countries worldwide <sup>65</sup>	Children aged 5–16 <sup>66</sup>	Math <sup>67</sup>
		MICS <sup>68</sup>	118 countries worldwide <sup>69</sup>	Children aged 5–17 <sup>70</sup>	Reading and math <sup>71</sup>
		UWEZO <sup>72</sup>	Kenya, Tanzania, and Uganda <sup>73</sup>	Children aged 6–16 <sup>74</sup>	Basic reading and math <sup>75</sup>

## NATIONAL EXAMS

*Countries should consider creating structures in their national examinations that can effectively support instructional improvement.* National exams tend to be paper-and-pencil- and census-based exams and are high stakes for students and schools, often determining placement for students or serving as proxies for school quality, resulting in school rankings in many countries. Their high-stakes nature makes them ripe for cheating, including by pushing low-performing learners out of the system to inflate outcomes,<sup>76</sup> and can act to narrow the focus of classroom instruction. To combat these challenges, **national exams should cover the breadth of the curriculum, including higher-order skills if possible, such as problem-solving, reasoning, and reflection; be used for incentives (based on improvement or models that control for other factors affecting learning outcomes) rather than punishments; and not be used to rank schools publicly if those rankings induce cheating or punitive measures.** It is important to ensure the wide and targeted dissemination of results, but in a way that spurs action rather than blaming.<sup>77</sup> For example, the Kenya National Examinations Council distributes results that show item-by-item correct and incorrect sample responses and common errors. Schools and teachers can use this information to inform instruction. It is important that FLN programs align with what is assessed in these national examinations given how powerful they are. In many cases, the skills targeted in FLN are building blocks for the skills tested in the exams—in this case, building awareness among stakeholders about this link may be helpful and may allow for more effective implementation of the FLN program.



## NATIONAL AND SUBNATIONAL ASSESSMENTS

Countries should consider national or subnational assessments to identify the portions of the country that require additional support, as well as particular skills that teachers need to improve instructionally. These are low-stakes, sample-, and typically paper-and-pencil-based assessments designed to align with the curriculum. They can be administered nationally or at mid-level administrative units within the education system and are used to examine the effectiveness of the system or the impacts of national or subnational education policies. The findings are often taken very seriously because they are generally owned and implemented by the government. However, national assessments tend to be given toward the end of primary school or in secondary school, which often comes too late to identify gaps in FLN learning outcomes and intervene accordingly. Learners who are behind in grade 1 and 2 tend to stay behind, and the gap between these learners and the rest of their class widens.<sup>78</sup> When implementing a national or subnational assessment, ensure that it includes early-grade assessment; is appropriately aligned within the education system (see Figure 6 on page 10); and includes benchmarks and targets. Consider using methodologies that allow for comparisons of results over time and across subnational contexts. See Box 4 on page 11 on benchmarks and linking of assessments.

## EARLY-GRADE READING AND EARLY-GRADE MATH ASSESSMENTS (EGRAS AND EGMAS)

Countries should consider using early-grade reading and mathematics assessments—such as EGRAs and EGMAs, or similar assessments—to identify the particular FLN skill levels of children and pedagogical challenges in the system. These assessments are school-based, sample-based, one-on-one FLN assessments usually administered for students in grades 1–3. They can be implemented at any level—from the national to the school level—and they use open-source tools that are adapted to the local context and language of assessment. The World Bank, USAID, and other institutions identified a gap in assessments in 2006, noting that national assessments and exams were usually not given until grade 4—too late to identify key learning gaps, as discussed above. Further, results of these assessments were often poor, and it was difficult to know whether that was because learners could not read and comprehend assessment instructions or because they did not have the knowledge being assessed. As such, in 2006, the World Bank, USAID, and others funded the creation of a simple and effective assessment of student reading outcomes in early grades—the EGRA. An example of the EGRA reading comprehension module is shown in Figure 2. Following the initial demand for EGRA, USAID funded the creation of EGMA for use in assessing early math outcomes in 2008.

EGRAs and EGMAs, and similar assessments, were designed to provide a system-level diagnosis of gaps in learning outcomes. Several countries, including Angola and Ghana, have adopted these tools as their national assessments. Angola established an assessment unit in the Ministry of Education and a dedicated US\$1 million budget line item<sup>80</sup> for these assessment activities. The ministry uses results from the assessment to inform in-service teacher training.

There have also been efforts to develop group-administered assessments that target many of the same skills addressed through EGRA. For example, the more traditional Group Administered Literacy Assessment (GALA) was designed to take advantage of lessons learned from the EGRA and assess a range of early-grade reading skills, while reducing administration time and simplifying scoring.<sup>81</sup> GALA has been used in Egypt, Ghana, Ethiopia, and

FIGURE 2. EGRA reading comprehension module example<sup>79</sup>

**Subtask 2b. Reading Comprehension**

When 60 seconds are up or if the child finished reading the passage in less than 60 seconds, tell the child you are going to ask them some questions about the story but first you want to allow them more time to re-read the story to themselves. Allow the student up to 3 minutes to read the story, and then ask if the student is done. If the student says yes, ask the questions after saying:

**Now I will ask you a few questions about the story you just read. Try to answer the questions as well as you can.**

**Kati njenda kubuuzayo ebibuuzo bitonotono ku ebyo by'osomye. Gezaako okuddamu obulungi nga bw'osobola.**

- Give the child at most 15 seconds to begin to answer the question, mark the child's response, and move to the next question.
- Read the questions for each line up to the bracket showing where the child stopped reading.

Sam lives on a farm. He has goats. Sam likes the goats. They are funny. They eat anything! He feeds the goats after school. Sam puts his homework near the goats. He plays football. Then his sister calls him. His homework is gone! A goat is eating something white. Now he knows where his homework is.

	Correct	Incorrect	No Response
<b>Where does Sam live?</b> [on a farm]			
<b>What does Sam have?</b> [goats]			
<b>What did Sam play?</b> [football]			
<b>Who calls Sam?</b> [his sister]			
<b>What happened to Sam's homework?</b> [a goat ate it]			



## BOX 2. Updating the EGRA

In recent years, EGRA users have made some adjustments to better understand students' abilities. For example:

- Providing extended time for the oral reading passage or an additional passage. This increases the number of children who are assessed on all of the corresponding comprehension questions.
- Adding an expressive language module, still under development by USAID in 2021. The module helps users understand the abilities of learners who receive zero scores on decoding and reading comprehension modules.
- Expanding the reading comprehension measures. This includes silent reading, untimed reading, maze, cloze, look back, and a variety of other ways to increase the reliability and validity of the EGRA comprehension measures.
- Developing and piloting other measures of spelling and oral language.<sup>83</sup>

Tanzania to save costs, while a similar tool, the Classroom-Based Early Grade Reading Assessment, was developed in Nepal, based on the GALA.<sup>82</sup>

EGRAs and EGMAs are most useful when they are integrated into government systems to maximize utilization, which may be easier to accomplish with group-administered versions. These tools have been criticized for having high per student costs.

## COMMUNITY-BASED ASSESSMENTS

Countries should encourage the implementation of community-based assessments in order to have disaggregated understanding of learning outcomes. Community-based assessments are conducted orally at the household level by community volunteers. Usually quick and easy to administer, these assessments capture FLN outcomes for both in- and out-of-school children of a wide age range. Examples of items from community assessments are included in Figures 3 and 4 below. While community-based assessments have proven successful in generating broad awareness of poor learning outcomes in many countries and also serve as a check on government assessment systems, they have seldom immediately resulted in changes to education systems as a whole. In the state of Bihar, India, however, they were used to identify high-needs communities, where the state government then implemented remedial instruction programs and created libraries, with positive effects.<sup>86</sup> When implementing a community-based assessment, ensure that there are champions at the levels of the system where results will be used and that those champions have actionable recommendations to actually improve learning.

FIGURE 3. Example UWEZO math assessment items<sup>84</sup>

**Sample of Mathematics test**

**Counting:**  
How many members are there in each set?  
(Should attempt any 5, atleast 4 must be correct)

**Number Recognition 10 – 99 :** (Should attempt any 5, atleast 4 must be correct)

15	47	25	23
30	94	36	51

**Number Recognition 100 – 999:** (Should attempt any 5, atleast 4 must be correct)

104	129	200	476
374	234	581	943

FIGURE 4. ASER reading assessment module<sup>85</sup>

**A big tree stood in a garden. It was alone and lonely. One day a bird came and sat on it. The bird held a seed in its beak. It dropped the seed near the tree. A small plant grew there. Soon there was another tree. The big tree was happy.**

**Rani likes her school. Her class is in a big room. Rani has a bag and a book. She also has a pen.**

e	d	w
s	c	
g	h	z
i	q	

hand	star
bus	
cat	book
day	few
	old
sing	bold

## SCHOOL-LEVEL ASSESSMENTS

School-level assessments consist of individual school assessments, local examinations, and formal inspections.

- **Individual school assessments** are conducted school-wide to inform school improvement plans. Head teachers often share results with teachers, parents, and school management communities, who then participate in school improvement planning.



- **Local examinations** are examinations set by a small number of schools to allow for a review and comparison of learning outcomes.
- **Inspections** involve visits by education officials, typically from the subnational level, to assess the quality or performance of schools and teachers through classroom observation, interviews with teachers, and the review of key documents and data.<sup>87</sup> Inspections may also be referred to as monitoring. It is important to differentiate inspections, which are intended to assess quality, from coaching or mentoring, which is not a type of assessment and is meant to provide support to teachers.

School-level assessments could be captured in a country's EMIS to allow local officials to follow up with schools to ensure that improvement plans are implemented and to allow local and national officials to target support to schools as needed.

### Uses of School-Level Assessments

While, historically, inspections did not include assessments of student learning outcomes, a growing number of countries are now including testing of small samples of students in their inspection visits to more clearly link school-level inputs (such as access to materials, school leadership, and instruction) with student learning outcomes. Some countries have also disseminated feedback from large-scale learning assessments through inspectors or coaches to inform adjustments to program elements at the school level. Results from local examinations can be used in the same way and are often a good opportunity to share and discuss learning results with parents. In some countries, local organizations such as UWEZO and Pratham have also played a role in tracking and disseminating learning outcomes with schools and communities to support school improvement. Examples of how student assessment results have been used at school-level include:

- **Engaging parents and the community in school assessments and improvement planning:** The Campaign for Female Education, which works in Ghana, Malawi, Tanzania, Zambia, and Zimbabwe, supports its alumni in returning to rural communities and serving as Learner Guides to young girls. The Learner Guides use tablets to collect data on learner retention, progress, and learning outcomes. They then share this information with the schools and communities to inform action planning. In return for their commitment, the alumni gain access to interest-free loans to grow their own businesses or receive a vocational teaching qualification that allows their fast-tracking into teacher training colleges.<sup>88</sup>
- **Embedding spot assessments of student learning outcomes in regular inspection visits:** Governments in Ghana, Jordan, Kenya, Rwanda, and Uganda have embedded assessments of a small sample of students in their inspection visits to better target areas for school support and teacher coaching. For example, the USAID Soma Umenye Activity worked with the government of Rwanda to incorporate grade 1 to 3 reading fluency and comprehension assessments into school inspections. Soma Umenye staff trained district education officers on how to conduct the assessments during school visits. Officers conduct the assessments using passages from that week's or the previous week's textbook lessons with five randomly selected students per class. The results of the assessments were then used to inform coaching conversations with teachers and discussions with head teachers. For example, if an assessment showed that students were decoding but not comprehending reading passages, support could then be directed to help teachers improve the quality of their comprehension strategies or to build more time for reading practice into the timetable. This monitoring activity has proven to be practical and cost-effective for governments.<sup>90</sup>
- **Sharing results from local examinations and large-scale learning assessments with schools:** School inspections can be leveraged to share results from local examinations and large-scale learning assessments with

#### BOX 3. Group-administered EGRA in Nepal

With support from USAID's Early-Grade Reading Program, the Ministry of Education, Science and Technology in Nepal developed a group-administered early-grade reading assessment, which it called CB-EGRA. This assessment was administered by teachers in all schools in the districts participating in the government's National Early Grade Reading Program. The results were shared at parent teacher association and village education committee meetings, where parent and community members discussed results and explored the reasons behind low reading comprehension achievement. These initiatives led all stakeholders, including local governments, school management committees, parent teacher associations, parents, and communities, to focus on ways to support children's academic achievement.<sup>89</sup>



teachers and head teachers and to coach them on how to use the results to make school-level improvements.<sup>91</sup> For example, the government of Malawi, with support from UNICEF and USAID's Yesani Ophunzira Activity (YESA), built on existing EMIS structures to include the following:

- Classroom observation data from Ministry of Education officials at the zonal level
- Data from school inspections conducted by district (the administrative unit above the zonal level) education officers
- Data on student reading outcomes from the country's National Reading Assessment

**This integration of student reading outcomes is a powerful addition to typical EMIS activities.** Primary education advisors and school inspectors enter data obtained from their school visits directly into the EMIS using a mobile application, accessible offline, that syncs with the database once officers have internet access. This information is then made available via a mobile dashboard to teachers, head teachers, coaches, inspectors, and education officials at all levels of the education system. **Even more powerful is the ability to use the data to inform the design and implementation of teacher training and coaching at the school level and the targeting of resources and support at the zone, district, and national level.**

### Best Practices for School-Level Assessments

To ensure that inspections and local examinations are effective:

- **Identify their purpose**, recognizing that school-level student assessments are most useful when they result in targeted support to teachers and head teachers.
- **Target support**: Inspect all schools and offer ongoing support to all teachers, but provide more frequent inspections and support to struggling schools (where assessment data suggest students are falling behind) and teachers.<sup>92</sup>
- **Ensure that inspection visits result in actionable feedback** with a sufficient level of detail to inform school and teacher adaptations.
- **Hold schools accountable**: While inspections and local examinations should not be high stakes (e.g., resulting in penalties for students, teachers, or schools), schools should be held accountable for delivering on clear recommendations. Coaches should help schools follow up on recommendations, and inspectors should follow up with schools to ensure that the recommendations are implemented.

### ASSESSMENT IN THE CLASSROOM

Assessments used in the classroom are teacher-administered formative and summative assessments of all learners that teachers can use to inform differentiated instruction and changes to their overall pedagogical approach or content focus. These assessments are discussed in more detail in the [Assessment-Informed Instruction: Classroom Level](#) guide and in a piece published by the [Global Reading Network on Formative Assessment](#). The aim is to ensure that teachers have and are able to use tools and methods to track learning progress and outcomes in their classrooms. This should be done without creating significant additional burden to teachers while also having a clear set of suggestions on how to improve instruction based on the information gained through assessment.

## Cost of Assessments

While inspection costs are usually covered by local governments, limited government resources often mean that they are not completed as frequently as needed to maximize impact. Conversely, costs for LSLAs are frequently borne by donors. Therefore, when considering costs, ensure that the potential benefits of the assessment outweigh the costs and that the assessment is sustainable beyond the FLN program lifespan.<sup>93</sup> Plan for the transition of costs to the government over time.

Costs vary significantly depending on the context, so it is important to examine the relevant costs of the different types of assessments in the system, as well as how the results from each assessment are currently being used. Some trade-offs to consider when trying to reduce costs include the following:

- **The comprehensiveness of assessments**: Assessments that will be used for monitoring the quality of the education system and targeting support to broad groups do not need to be as lengthy or comprehensive (e.g., they might assess only a few key skills as opposed to the entire curriculum) as assessments focused on targeting

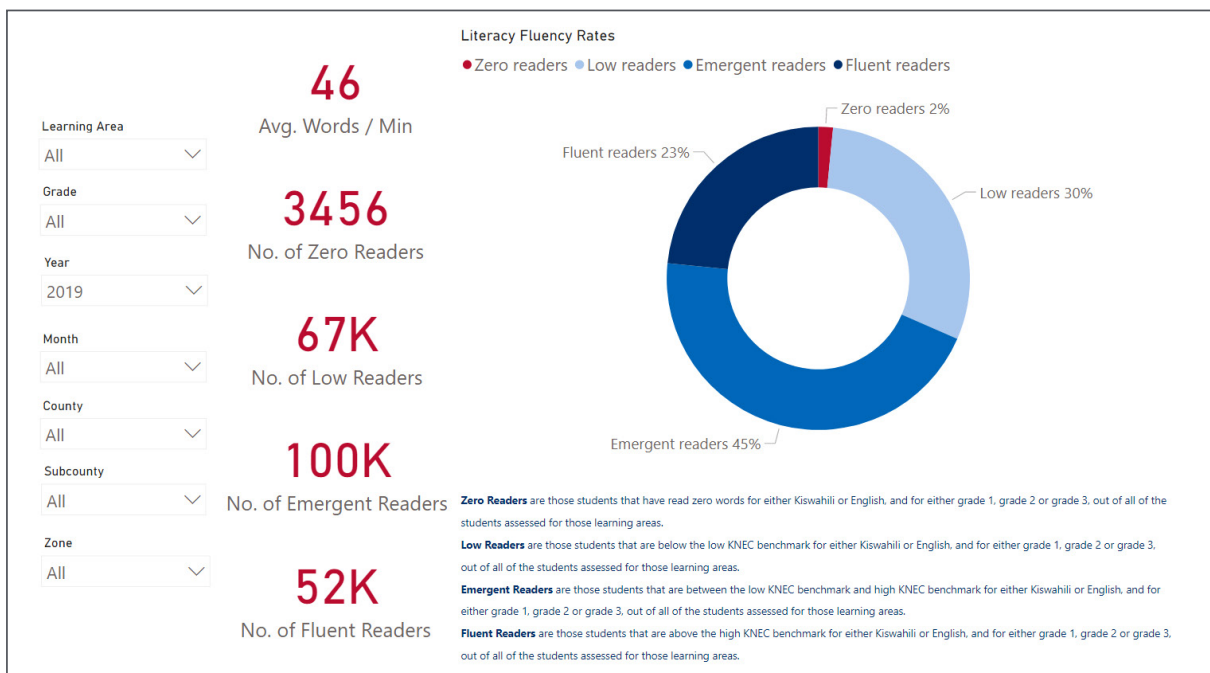




support to individual teachers or students, which should cover the breadth of the curriculum to help diagnose specific gaps in learning.

- Size of the sample:** Using smaller samples, including lot quality assurance sampling (LQAS)<sup>94</sup> methods, can be useful for identifying broad trends in learning outcomes that might suggest system weaknesses. For example, if an LQAS assessment conducted by an inspector of five students in a classroom finds that all sampled students are able to correctly multiply numbers, then it can be inferred that most students are likely able to correctly multiply numbers and that the teacher does not need support in teaching multiplication. On the other end of the spectrum, assessments that will be used to diagnose learning gaps or advance students to the next level require census-based outcome measures.
- The use of local government officials or centrally trained enumerators to gather data:** Using local enumerators or government officials to carry out assessments can reduce travel costs. If this is done, be sure not to assign officials to schools that they already oversee, to avoid any potential conflicts of interest that may bias results.
- The complexity of information required:** Increasing the time spent on individual data collection will obviously increase costs. Some of the time spent might be essential, such as measuring students' comprehension and oral reading fluency using EGRA. But consider the marginal value added from such things as adding more students to the sample or more items to the instrument. More students may be needed only if the assessment will be used for promotion of differentiated instruction purposes, or if the small sample size does not result in any noticeable trends. More items may be necessary if outcomes suggest a gap in a key skill but there is not enough information on why (e.g., in many countries in sub-Saharan Africa, learners struggle with reading comprehension). It may be that this is a result of issues with language comprehension, vocabulary, or another pre-cursor skill; adding items that measure this knowledge might help education officials better understand system weaknesses.
- Rapid inexpensive dashboards:** Developing a large-scale assessment system provides an opportunity for data to be rapidly and inexpensively shared to support accountability and decision-making. For example, in Kenya, the Ministry of Education is expanding on the utilization of a dashboard that was introduced under the Tusome program (see Figure 5), which includes student assessment data disaggregated by Kenyan benchmarks and data from classroom visits. This dashboard is accessible by education officials at the national level, county level, and local level. Such virtual dissemination of data has the benefit of incurring only marginal costs. Moreover, studies suggest that the Tusome dashboard tool has stimulated a culture of data use within the Kenyan government more generally.<sup>95</sup>

**FIGURE 5. Kenyan government data dashboard example<sup>96</sup>**





## IMPROVING SYSTEM-LEVEL ASSESSMENT

Based on an understanding of the importance of and different types of assessments available for use in your country, consider ways in which the assessment system in your context may be able to improve:

### 1 EMBED ASSESSMENT INTO THE EDUCATION SYSTEM

Donors and partners often design assessment systems that are not embedded into the education system from the start, which negatively affects government buy-in, use of results, and sustainability. To avoid this mistake, donors and partners, in collaboration with the government, should either work to improve existing assessments or, where assessments do not exist, legislate the establishment, conduct, and financing of assessment programs and agencies, including identifying which agency will conduct the assessments and which will use the results. For example, when the government of India, together with partners, identified the need for a census-based school assessment to diagnose learning gaps, it legislated the creation of one for grades 3, 5, and 8 in its 2020 National Education Policy.<sup>97</sup> The Central Square Foundation then worked with the government's Central Board of Secondary Education (CBSE) to develop the assessment and ensure alignment with the curricula and the Global Proficiency Framework (see Box 4 below on setting benchmarks). CBSE will implement the assessment, and the results will be used at the school level to support differentiated instruction. The data may also be used at the state and national level to target support to schools, as well as disseminated publicly (in anonymized form).

Despite the promise of India's new assessment, in other settings, often the most effective thing one can do is not to roll out new assessments but to work with the government to make adjustments to the dissemination of results from existing assessments and connect that to program implementation and instruction. Funding considerations should also include setting aside funds for any interventions that result from the use of assessment findings. When working with the government, ensure an inclusive process that involves the participation of key policy makers and practitioners (including teachers) in each stage of assessment design, data collection, analysis, and dissemination of results. Including government officials in the process tends to reduce costs when they replace external hired support. For example, in Malawi, USAID's YESA<sup>98</sup> activity supports the Ministry of Education, Science, and Technology's National Reading Assessment and uses district education managers, primary education advisors, and other government officials to collect the data (rather than hired enumerators), and ministry staff are currently being trained to analyze and disseminate results.

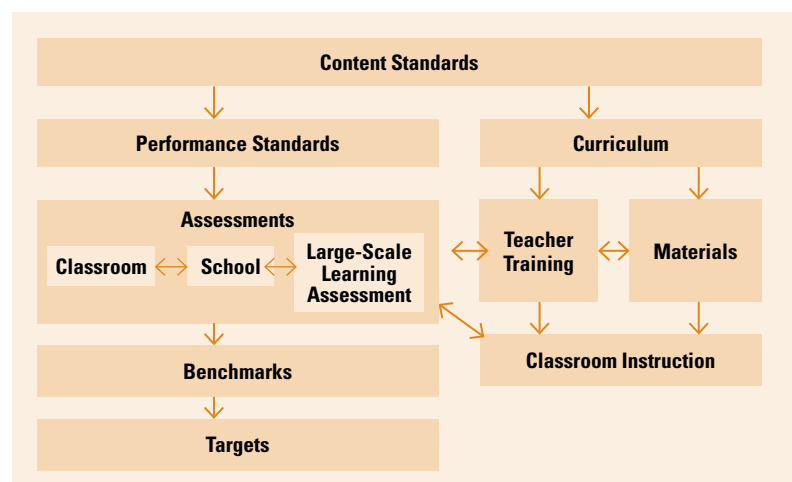
### 2 BE REALISTIC ABOUT COSTS AND SUSTAINABILITY

Financial constraints can hinder the sustainability of assessments. Be sure to consider the cost trade-offs discussed earlier in this guide and ensure that decisions are made with a clear understanding of whether the government will be able to sustain those costs over time. Consider gradually transitioning donor-funded assessments to the government over the life of an FLN program.

### 3 ENSURE ALIGNMENT OF ASSESSMENTS WITHIN THE EDUCATION SYSTEM

When designing or reviewing an education system in coordination with key stakeholders, it is important to ensure the alignment of assessments with key components of the system. This key element is also discussed in the Assessment-Informed Instruction: Classroom Level guide. As Figure 6 illustrates, there needs to be coherence between educational standards, benchmarks, and assessments, and these all need to be closely linked to curricula, teacher training, and materials. Tools such as the Surveys of Enacted Curriculum<sup>99</sup>—which

FIGURE 6. Aligning assessment within the education system





examine the relationship between curricula, instruction, and assessment—can help identify how well an assessment system aligns with the education system and what gaps need to be addressed. While it may not always be possible to conduct a formal Surveys of Enacted Curriculum analysis, it is critical that your assessment efforts examine how a given assessment aligns with existing educational standards, curricula, and instructional practices. Alignment ensures that teachers stay focused on critical content and that assessment results reflect actual learning in the classroom, which is critical to ensuring assessment-informed instruction. System alignment analysis can be done through a review of curricula and standards documents; interviews and discussions with education officials, administrators, and teachers; and classroom observations.

Also consider alignment of the system assessment and curricula with standards for global minimum proficiency, as described in Box 4, if there is a desire to use the assessment to report on progress toward Sustainable Development Goal 4.1.1, which focuses on education quality.

#### BOX 4. Setting benchmarks using policy linking for measuring global learning outcomes

Policy linking is a process that links assessments to content and performance standards and sets benchmarks for minimum proficiency standards. The methodology engages master teachers and curriculum specialists to act as panelists in workshops where they make informed judgments about how minimally proficient students would likely perform on a given assessment. Government stakeholders have reported improved ownership over benchmarks set using policy linking methods.

Once benchmarks have been established through policy linking, new benchmarks for future assessments can be set using policy linking or statistical linking methods (see the box on developing equivalent instruments below).

Policy linking can be used in conjunction with the [Global Proficiency Framework](#), which defines a set of common global content and performance standards adopted by the UNESCO Institute for Statistics, to set benchmarks that link assessments to Sustainable Development Goal 4.1.1, which focuses on education quality.<sup>101</sup> Countries may also use statistical methods such as LAYS to link their national or subnational assessments to a regional or international assessment that is already linked with SDG 4.1.1.

#### 4 CLARIFY THE PURPOSE OF ASSESSMENTS

Another key mistake that donors, governments, and partners often make is conducting assessments for assessment's sake rather than to inform targeting of support and to improve instruction. Early in the design process, work with the government, donors, partners, and other key stakeholders to identify the purpose of the assessments. When doing so, it might be helpful to keep in mind the three core functions of the education system, as identified by Luis Crouch and Joe DeStefano:<sup>100</sup>

- **Set expectations for learning outcomes:** It is important for education systems to have well-defined content and performance standards, as well as benchmarks and targets. Without agreed-upon goals, donors, governments, partners, and teachers do not know where they are going or how to allocate resources to navigate there. If the education system does not have benchmarks and targets to help you navigate, work with key stakeholders to plan benchmarking and target-setting workshops and ensure wide dissemination of the goals to teachers, communities, and other key stakeholders.

Benchmarks should identify the score or performance level that a student should reach to be considered proficient in a particular skill (e.g., at end of grade 2, students should be able to achieve a score of 50 out of 75 on the National Reading Assessment). There are several ways of setting benchmarks. For example, the UNESCO Institute for Statistics and USAID have developed policy linking for measuring global learning outcomes, described in Box 4 (above). This complements UNESCO's existing work to statistically link learning assessments to increase comparability of results across countries and across different assessments.

Targets should provide countries with intermediate, realistic goals for the percentage of a given population that should meet the benchmark in a given period of time (e.g., by the end of 2021, 35% of grade 2 students should meet the benchmark). Targets are best set by first considering current student performance and estimating the rate of change that is realistic yet pushes the system toward improvement. The current rate of improvement in countries is too slow to ensure achievement of global goals that all children attain minimum proficiency in reading and math by 2030. It is important that countries seek ways to accelerate improvement using assessment-informed instruction and that this goal be reflected in countries' targets—with sufficient resources devoted to ensuring that the goal is attainable.



**Universal design for assessment (UDA) approaches can help ministries reach their benchmarks by ensuring that assessments used are inclusive and accessible for all (see Box 5 on UDA).**

Assessments should be designed with the UDA framework in mind and in close coordination with key stakeholders.

**Monitor and hold schools accountable for meeting those expectations:**

The government needs to monitor the overall performance of the education system with respect to standards, benchmarks, and targets. System-level monitoring can be accomplished through periodic census- or sample-based assessments of student learning outcomes,<sup>103</sup> which can be achieved through any LSLA or LQAS. That said, national exams, when well designed, are the best tools for holding students and teachers accountable since they tend to use census-based samples. At the same time, community assessments (such as those used by ASER and UWEZO) have the advantage of giving instantaneous feedback to parents and the community about learning outcomes. However, as described above, monitoring is useful only if it spurs action—in the form of resources and support for schools not meeting expectations.

**Intervene to support the students**

**and schools that are struggling and ensure equity for all students:** National, subnational, non-standardized, and citizen-led assessments—when they use sampling methodologies that allow for the representativeness of the groups of interest and collect relevant demographic information—can be used to target support to the groups most at need. To identify the schools most at need, a large sample is necessary. It is also important to ensure that assessments themselves are inclusive, as otherwise, large groups of students and schools in need of support may be missed (see Box 5). Teaching at the Right Level (TaRL) methods also help ensure equity for all students. The TaRL approach uses learning assessment in the system to identify students for grouping by their performance level and then tailors instruction to the level of each respective group. TaRL typically demands significant teacher training and ongoing support to implement, but has been used successfully in India and across ten African countries.<sup>104</sup>

### BOX 5. Universal design for assessment<sup>102</sup>

When designing or refining assessments, it is critical to ensure inclusivity. Assessments should be designed to be universal—accessible by all—from the beginning. This means:

- **Removing barriers to assessment**—For example, remove cognitive, sensory, emotional, and physical barriers (such as asking learners to point, while not providing an option for those who cannot do so).
- **Building accessible items from the beginning**—For example, do not reference colors, which are not accessible to learners who are blind, or music that is not accessible to learners who are deaf.
- **Designing with accommodations in mind**—Consider physical, intellectual, and learning disabilities.
- **Using clear, simple instructions and language**

**UDA Accommodations for Receiving Information**

- Have someone read text out loud or allow audio recordings.
- Display fewer items on a page or use large print or braille.
- Give learners a written list of instructions.

**UDA Accommodations for Expressing Information**

- Allow oral responses or use of alternative and augmentative communication devices or sign language.
- Allow use of a calculator or math tables.
- Allow written responses in braille.

**UDA Accommodations for Motivation**

- Provide additional time to complete the test.
- Permit test taking to occur over several sessions, allowing for breaks.
- Allow test taking at times of the day when a specific learner is most focused.
- Allow test taking in alternative settings with fewer learners and distractions.

## 5 FOCUS ASSESSMENTS ON WHAT TEACHERS AND POLICY MAKERS NEED TO KNOW

**Focus the assessment's design on collecting only the information that is necessary to improve teaching and, thereby, learning outcomes.** A few ideas on how to simplify assessments, increase their effectiveness, and reduce costs, while ensuring quality, include the following:

- **Measure important skills:** Ensure that the skills targeted in the assessment align with content and performance standards. If an education system hopes to develop higher-order skills (e.g., comprehension, critical thinking, and problem solving), assessments in the system need to include at least some measure of those skills, especially since assessments often drive what is taught in class. While it is easy to design items that require automaticity or recall (such as naming letter-sounds or doing a simple



addition problem), it is more difficult to design items that require comprehension, critical thinking, and integration of skills, which is why many LSLAs either do not measure higher-order skills at all or fail to effectively measure them. Higher-order skills are often best assessed through open-response items rather than closed-response items (e.g., true-false and multiple-choice options).

• **Ensure the tools measure the intended skills:** Consider the use of oral items (such as those used in EGRAs, EGMAAs, and community-based assessments) in assessments of FLN so that reading ability is not conflated with math ability, and decoding ability is not conflated with comprehension. For instance, if responding to items in a written test, children may get an item wrong because they could not read the instructions, even if they would otherwise have been able to answer correctly. Take, for example, a math word problem that requires a learner to read the problem and then solve it. If the child gets the problem incorrect, it is impossible to know whether that is due to a gap in the child's reading skills or math skills. Similarly, questions should not require unnecessary background information that may not reflect typical cultural knowledge and classroom instruction (e.g., the need to know that snow is white in a context where it never snows or to be familiar with a food or animal not common in the context).

• **Be appropriately leveled to the grade and ability of students:** Very few students should get all items right or all items wrong on an assessment, as this limits the useability of assessment information. To avoid this problem, engage teachers in helping develop or refine tests, and pilot test instruments to see if the assessment has ceiling or floor effects. For example, if EGRAs and EGMAAs being used in a pilot are finding a high number of zero scores, more basic subsections may need to be added, and the dropping of more advanced subsections might be considered—or a stop rule applied where children are asked not to continue to the next section if they have zero scores on previous sections.

• **Include the right number of items for the purpose:** Long assessments are tiring for students and can result in decreased performance. Use only the number of items needed for the purpose. Assessments used to promote students or differentiate them into ability groups may have more items than assessments focused on monitoring the effectiveness of the education system as a whole. Piloting will help determine whether the number of items needs to be changed.

• **Include contextual information that will actually be used:** Policy makers need more data than just learning outcomes (see Box 7). However, many assessment systems collect contextual information that is never used by policy makers or other stakeholders. To reduce costs, it is thus important to carefully consider which additional data to gather with an assessment. Some information that has proven useful includes (1) languages spoken by the children and teacher (see the [Practical Language Choices guide](#) for more discussion of this and other

### BOX 6. Developing equivalent instruments<sup>105</sup>

To ensure stakeholder confidence in learning assessment outcomes and in comparing outcomes over time, you should consider the following:

**One or multiple instruments:** Instruments should be changed from year to year only if there are concerns of tests leaking. If not, the same assessment can be used over time, assuming that it is still contextually relevant. If multiple versions of the assessment are needed, make sure they are at the same level of difficulty.

**Ensuring similar difficulty:** Items that focus on naming letters or numbers can remain the same, with the order scrambled. Reading passages can be modified by simply changing names, story subjects, actions, and adjectives with grade-level equivalents. Parameters or criteria can be used to ensure that math problems can be adjusted without changing the level of difficulty.

**Equating:** Despite creating instruments to be as similar as possible, there will inevitably be small differences in difficulty. These differences can be addressed by either having a sample of students who are representative of the country take both assessments or including common anchor items, and then using that information to make results equivalent.

### BOX 7. Disaggregating assessment results<sup>106</sup>

Assessment in the system data should be disaggregated by factors such as:

- Sex
- Ethnicity
- Geographical region
- Urban versus rural schools
- Learners with disabilities (this information should be informed by a rigorous screening process)
- Other key characteristics specific to the context

This information should be used for targeting support to schools and students based on assessment outcomes by population, as well as for considering whether the curriculum and teaching approaches are inclusive of all groups.



language-of-instruction issues); (2) whether the teacher has received training; (3) whether the children have the textbook or workbook; (4) whether the teacher has and uses a teacher's guide; (5) and whether teachers implement the FLN program and instructional practices introduced through training in the classroom.

- **Be comparable over time:** If assessment instruments are expected to change from year to year but you still plan to use them to compare results and track learning outcomes over time, you will need to create comparable assessments. See Box 6 above for tips in this regard.

## 6 DECIDE WHEN AND HOW OFTEN TO COLLECT ASSESSMENT DATA

The frequency and timing of assessments should be based on their purpose and available funding. School-level assessments, meant to inform and target support, should occur several times a year. Assessments focused on monitoring system changes need occur only every two years. Assessments aimed at comparing outcomes between countries can occur even less frequently, at four- or five-year intervals. Assessments focused on monitoring outcomes and comparing them with benchmarks should occur at the same time each year, preferably at the end of the school year.<sup>107</sup>

## 7 FIND A BALANCE BETWEEN ACCOUNTABILITY AND LEARNING

Educational assessment must strike a fine balance: if there is no accountability associated with assessment results, assessments will do little to effect change; however, if the stakes are too high, then assessments can result in “teaching to the test,” test leakage, or teachers asking poor-performing students to stay home on test day.<sup>108</sup> **Assessment must promote improvement, whether through accountability or learning mechanisms; it should be used to help programs adapt.** To improve accountability for results and to reduce incentives to cheat, consider implementing the following measures:

- **In order to focus on learning, disseminate results at the school and teacher level,** targeting teachers and education officials with information that they can use to reform practices, rather than emphasizing school-ranking reports. Ghana and Kenya, for example, have reduced their use of such reports, effectively reducing many of the negative effects of their national exams. Ghana introduced ranking (or school league) tables in 2004, but that led to a sharp increase in malpractice. When it eliminated the use of these tables, cheating was reduced. A more limited release of district-level examination results at the end of the basic level of schooling has been associated with a positive impact in a few districts.<sup>109</sup> There must be a careful balance between accountability for results and creating incentives for fraud. In some countries, having access to publicly available data increases accountability, but in others, it can induce negative behaviors. Understanding whether the available results are helping or hurting is important.
- **Provide support to and then hold schools and teachers accountable for taking actions** based on assessment results, through the use of coaching, other ongoing support, and inspectors.
- **Use incentives for improvement rather than penalties** (and consider factors such as students' socioeconomic background, school resources, and so forth, that may be affecting scores, using assessment to identify ways of meeting the needs of populations that are disproportionately affected by these challenges; the idea is to ensure that the system meets the needs of even the most disadvantaged schools and populations). For example, providing certificates and recognition when scores improve, or providing additional resources to help struggling schools, can be effective in triggering improvement.

### BOX 8. Disseminating school-level results in India<sup>110</sup>

In Maharashtra State in India, all schools complete a school evaluation that includes both learner assessments and other contextual assessment components. The schools then gather together head teachers, teachers, students, parents, and members of the school management committee to facilitate school improvement planning. The school management committees produce evaluation reports (including a list of areas for improvement) via a centralized digital dashboard, which is consolidated and used by officials at the cluster, block, district, state, and national level to identify school-specific needs and common areas of intervention that might improve school performance. School inspectors follow up on the improvement plans to track progress during their visits.

## 8 ENSURE TARGETED DISSEMINATION

**Effective dissemination is one of the most important aspects of a high-quality assessment system.** Yet the dissemination of assessment results, particularly from LSLAs, is often not done effectively. Some suggestions for ensuring targeted dissemination include the following:



**School administration and teachers:** If large-scale assessments are going to maximize their benefits, school administrators and teachers need specific information about learner outcomes, which items students tended to get right/wrong, what the most common incorrect response was, and suggested ways to address them, as shown in Box 9.<sup>111</sup> Ongoing support to teachers in the implementation of those actions is paramount. In addition, data should be timely, reaching schools and teachers not too long (within a few months) after an assessment has ended. This can be accomplished through the use of live dashboards for disseminating results from systematic assessments. The Jordan RAMP initiative, for example, successfully used government data collection tools to provide schools with school-specific data about student performance, which it complemented with school-level community meetings to discuss the results and suggest strategies to respond to instructional challenges by teachers and parents alike.<sup>112</sup>

**Decentralized education officers:** Like the other stakeholder groups, education officials at regional or district levels need data presented in concise, user-friendly, and timely formats. They may also need guidance in using the data to target support and resources to schools and teachers. The USAID Ghana Learning activity, for example, developed live dashboards that district and regional education officers reviewed twice per term and then used to inform district action plans to improve support for schools. The dashboards were created using Power Bi software and included head teacher- and teacher-entered school records, such as absentee rates, circuit supervisor information from school visits and classroom observations, and assessment data from PACE assessments. The result was simple displays that showed districts and schools in need of the most support (red), some support (yellow), or limited support (green), as shown in Figure 7 above. The team that created the dashboard recommends collaborating closely with local education officers to tailor dissemination and completing several rounds of user testing and feedback.<sup>113</sup> Discussions with education officers using the dashboard in Figure 7 would focus on encouraging teachers to continue the best practices identified in the dashboard, but to increase the number of coaching sessions per teacher and the school-based meetings held.

**BOX 9. Example of detailed feedback on student assessment performance**

Item 1:  $45 - 5 \times 4 =$

- a. 160    c. 25
- b. 20    d. 16

**Correct Response**

c. 25 (provided by 15% of students)

**Most Common Incorrect Response**

a. 160 (60%)

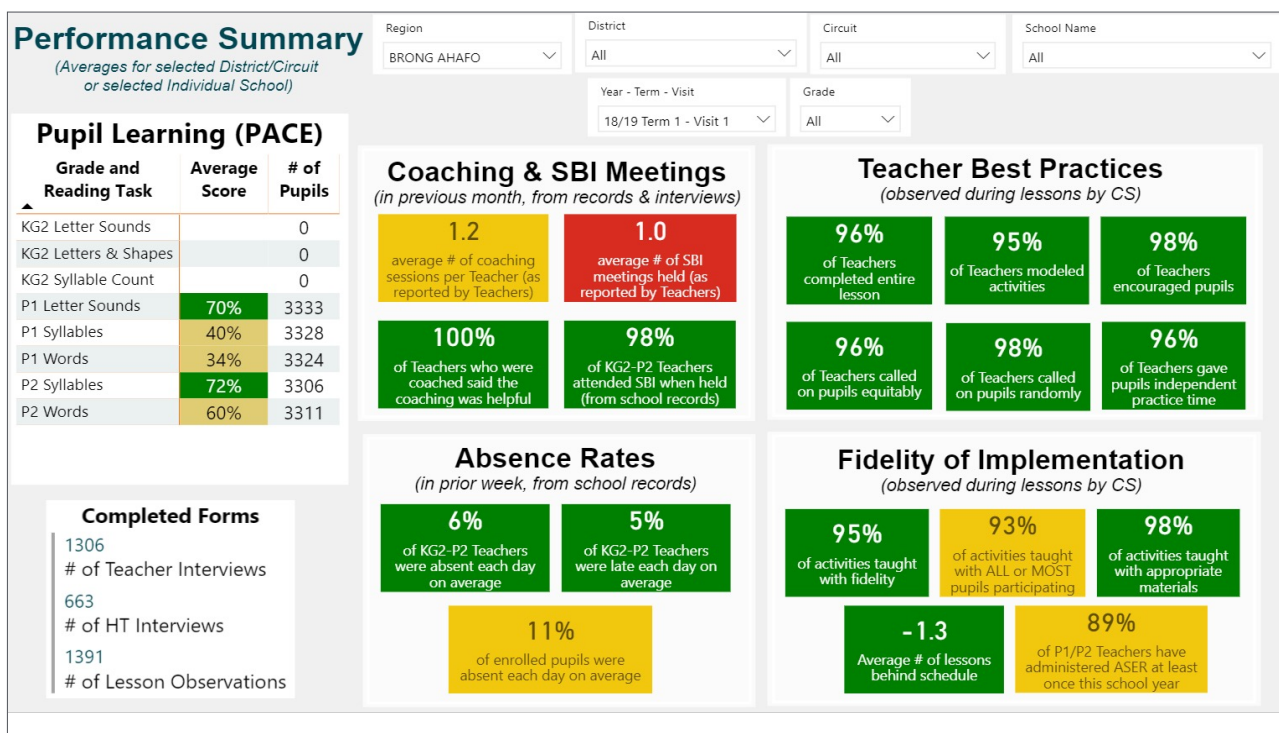
**Reason for Error**

Lack of mastery of the concept of order of operations

**Recommendation**

Review and practice order of operations with students

**FIGURE 7. Ghana Learning activity dashboard**





- **National policy makers:** In addition to using data to target their support to schools, national policy makers and education officials may also use assessment data to inform changes to curricula or education policies, or to set benchmarks and targets. Indeed, in a number of countries, the appropriateness of curricula has been challenged in the wake of findings from national assessments. For example, a Bhutanese national exam led the Ministry of Education to reform the country's math curriculum after it determined that students were experiencing curriculum overload; and the government of Ethiopia made changes to its curricula following feedback from teachers, students, and parents on the reasons behind the poor test results.<sup>114</sup>
- **Parents and the community:** Parents and communities are well positioned to have a strong voice when it comes to advocating for school improvements; so, it is important to identify ways to share results with them. In Nepal and India, village community meetings and information campaigns have been used as channels for disseminating assessment outcomes. Such campaigns have used community outreach to increase awareness of the roles and responsibilities of school oversight and management committees and to disseminate information on a repeated basis, such as through festivals, posters, paintings, take-home calendars, and learning assessment booklets. The parent and community involvement generated through these campaigns has driven improvements in learning outcomes. When disseminating information to communities, one touchpoint is likely insufficient, and key actions that the community can take should be an integral part of the message.<sup>115</sup>
- **Media:** The media can play a role in ensuring that assessment results are taken seriously by the government and stakeholders, putting pressure on the system to effect change. That said, you should prioritize ensuring that results get to the key users first (with details on specific actions that they can take), since disseminating results to the media alone will not result in change. Results should be shared in as many mediums as possible (e.g., television, radio, social media, WhatsApp) to reach a wide audience. The media can also be useful to celebrate and promote successes to incentivize other schools to improve, by acknowledging the most improved schools by region and other milestones. This is also a potential use of results for local and national education officials.

---

## Conclusion

Assessment enables schools, governments, and other stakeholders to understand how well students are learning and to identify where improvement is needed. An assessment system that is well aligned to curricular standards and benchmarks can provide information about what works to increase learning outcomes, how instruction might be improved, and how support can best be targeted. Armed with an understanding of the different types of assessment available and the way in which they should work together in a coherent system, and considering the recommendations laid out in this guide, you can be on the way to an assessment system that can better support the improvement of learning outcomes.





## RESOURCES

### Analyzing the assessment system

- World Bank's SABER: <https://saber.worldbank.org/index.cfm?indx=5&sub=2> (see the section on Student Assessment)
- GPE's ANLAS: <https://www.globalpartnership.org/sites/default/files/docs/2020-11-GPE-ANLAS-manual.pdf>

### Benchmarking

- Policy Linking for Measuring Global Learning Outcomes Toolkit: <https://www.edu-links.org/resources/toolkit-setting-internationally-linked-benchmarks-early-grade-reading-and-math>

### Education system alignment

- Surveys of Enacted Curriculum: <https://curriculumanalysis.org/products-SEC.asp#>

### Implementing a national learning assessment

- UNESCO's *Quick Guide No. 3: Implementing a National Learning Assessment*: <http://uis.unesco.org/sites/default/files/documents/quick-guide-3-implementing-national-learning-assessment.pdf>

### Implementing a national exam

- World Bank's *Public Examinations Examined*: <https://openknowledge.worldbank.org/bitstream/handle/10986/32352/9781464814181.pdf?sequence=2&isAllowed=y>

### Implementing an EGRA/EGMA

- EGRA Toolkit and EGRA Toolkit 2.0: [https://pdf.usaid.gov/pdf\\_docs/PA00M4TN.pdf](https://pdf.usaid.gov/pdf_docs/PA00M4TN.pdf) and [https://www.globalreadingnetwork.net/sites/default/files/media/file/EGMA\\_Toolkit\\_FINAL\\_0.pdf](https://www.globalreadingnetwork.net/sites/default/files/media/file/EGMA_Toolkit_FINAL_0.pdf)



### TECHNICAL EXPERTISE NEEDED

- Assessment development and analysis, to support development, equating, analysis of assessments, and benchmarking.
- Reading or mathematics education, with focus on early-grade content and instruction, to support assessment development in targeting key skills appropriately.



## ENDNOTES

- 1 UNESCO Institute for Statistics, "More Than One-Half of Children and Adolescents Are Not Learning Worldwide," Fact Sheet No. 46 (September 2017); United Nations, *The Sustainable Development Goals Report 2019* (New York: United Nations, 2019).
- 2 Marguerite Clark, "Measuring Learning: How Effective Student Assessment Systems Can Help Achieve Learning for All," Education Notes 67346 (World Bank, 2012); OECD, *PISA 2018 Results (Volume VI): Are Students Ready to Thrive in an Interconnected World?* (Paris: OECD, 2020).
- 3 Girindre Beeharry, "Accelerating Foundational Literacy and Numeracy through Focus, Measurement, Support and Accountability," *World Bank Blogs* (October 19, 2020).
- 4 Phil Elks, "The Impact of Assessment Results on Education Policy and Practice in East Africa," DFID Think Piece (January 2016).
- 5 John Cresswell, Ursula Schwantter, and Charlotte Waters, *A Review of International Large-Scale Assessments in Education: Assessing Component Skills and Collecting Contextual Data* (Paris: OECD, 2015).
- 6 Ieva Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," IIER-UNESCO Working Paper (April 2019); Ina V. S. Mullis, *TIMSS 2019 Assessment Frameworks: Introduction* (Amsterdam: IEA, 2017)
- 7 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019; OECD, *PISA 2018 Results*, 2020.
- 8 Ibid.
- 9 Ibid.
- 10 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 11 Ibid.
- 12 Ibid.
- 13 Ibid.
- 14 Ibid.; OECD, *PISA 2018 Results*, 2020; Clark, "Measuring Learning," 2012.
- 15 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019; Mullis, *TIMSS 2019 Assessment Frameworks*, 2017.
- 16 Ibid.
- 17 Ibid.
- 18 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019; OECD, *PISA 2018 Results*, 2020; Clark, "Measuring Learning," 2012.
- 19 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 20 Ibid.
- 21 Ibid.
- 22 Kate Anderson, "Strengthening Learning Assessment Systems," Knowledge and Innovation Exchange Discussion Paper (July 2019).
- 23 Ibid.
- 24 Ibid.
- 25 Ibid.
- 26 Australian Council for Educational Research, "The Latin-American Laboratory for Assessment of the Quality of Education: Measuring and Comparing Educational Quality in Latin America," Assessment GEMS Series, No. 3 (March 2014).
- 27 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019; OECD, *PISA 2018 Results*, 2020; Clark, "Measuring Learning," 2012.
- 28 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 29 Ibid.
- 30 Ibid.
- 31 Australian Council for Educational Research, "The Latin-American Laboratory for Assessment of the Quality of Education," 2014.
- 32 Ibid.
- 33 Ibid.
- 34 Ibid.
- 35 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019; OECD, *PISA 2018 Results*, 2020; Clark, "Measuring Learning," 2012.
- 36 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 37 Ibid.
- 38 Ibid.
- 39 Vincent Greaney and Thomas Kellaghan, *Assessing National Achievement Levels in Education* (Washington, DC: World Bank, 2008).
- 40 Anderson, "Strengthening Learning Assessment Systems," 2019.
- 41 Ibid.
- 42 Ibid.
- 43 A 2018 review of education in sub-Saharan Africa recommended the elimination of primary-school leaving exams given that they are a barrier to access to secondary schools.
- 44 Anderson, "Strengthening Learning Assessment Systems," 2019.
- 45 Thomas Kellaghan, Vincent Greaney, and T. Scott Murray, *Using the Results of a National Assessment of Educational Achievement* (Washington, DC: World Bank, 2009).
- 46 Subnational assessments are a variation on national assessments in which the sample is confined to a region, district, province, or state within a country. Such assessments are often carried out in large, decentralized systems with different education policies or languages of assessment. The assessments are generally not linked between different areas of a country, which makes direct comparisons of student achievement between regions difficult (Greaney and Kellaghan, *Assessing National Achievement Levels*, 2008).
- 47 Kellaghan, Greaney, and Murray, *Using the Results of a National Assessment of Educational Achievement*, 2009.
- 48 Ibid.
- 49 Ibid.
- 50 Ibid.
- 51 Amber Gove, *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition* (Washington, DC: USAID, 2015).
- 52 Ibid.; Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019; Thomas Kellaghan and Vincent Greaney, *Public Examinations Examined* (Washington, DC: World Bank, 2020).
- 53 Ibid.
- 54 Ibid.
- 55 Gove, *Early Grade Reading Assessment (EGRA) Toolkit*, 2015; Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 56 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 57 Ibid.
- 58 Ibid.
- 59 Kellaghan, Greaney, and Murray, *Using the Results of a National Assessment of Educational Achievement*, 2009; ASER Centre, *Annual Status of Education Report (Rural) 2018* (New Delhi: ASER Centre, 2019).



- 60 ASER Centre, *Annual Status of Education Report (Rural) 2018*, 2019.
- 61 Ibid.
- 62 Ibid.
- 63 Ibid.
- 64 PAL Network, *ICAN: International Common Assessment of Numeracy; Background, Features and Large-Scale Implementation* (Nairobi: People's Action for Learning Network, 2020).
- 65 Ibid.
- 66 Ibid.
- 67 Ibid.
- 68 UNICEF, "About MICS" (2019), <https://mics.unicef.org/about>.
- 69 Ibid.
- 70 Ibid.
- 71 Ibid.
- 72 Raudonyte, "Use of Learning Assessment Data in Education Policy-Making," 2019.
- 73 Ibid.
- 74 Ibid.
- 75 Ibid.
- 76 Jacobus Cilliers, Isaac M. Mbiti, and Andrew Zeitlin, "Can Public Rankings Improve School Performance? Evidence from a Nationwide Reform in Tanzania," *Journal of Human Resources* (March 2020).
- 77 Daniel Wagner, Marlene Lockheed, Ina Mulis, et al., "The Debate on Learning Assessments in Developing Countries," *Compare: A Journal of Comparative and International Education* 42, no. 3 (2012): 509–545.
- 78 ACDP Indonesia, "The Critical Importance of Early Grade Reading and Assessment" working paper (December 2014).
- 79 Adapted from Uganda Literacy Achievement and Retention Activity, EGRA and Supplementary Data Collection Instruments, <https://shared.rti.org/content/ugandalara-egra-and-supplementary-data-collection-instruments>.
- 80 RTI International, *Early Grade Mathematics Assessment (EGMA) Toolkit* (March 2014), [https://ierc-publicfiles.s3.amazonaws.com/public/re-sources/EGMA%20Toolkit\\_March2014.pdf](https://ierc-publicfiles.s3.amazonaws.com/public/re-sources/EGMA%20Toolkit_March2014.pdf).
- 81 RTI International, *An Addition to the Toolbox for Measuring Literacy Skills of the Youngest Students: The Group Administered Literacy Assessment (GALA)*, [https://ierc-publicfiles.s3.amazonaws.com/public/resources/1071\\_GALA\\_brief\\_02-16-17\\_r6.pdf](https://ierc-publicfiles.s3.amazonaws.com/public/resources/1071_GALA_brief_02-16-17_r6.pdf).
- 82 RTI International, *USAID's Early Grade Reading Program (EGRP) in Nepal: Final Progress Report, Years 1–5: March 2015–October 2020* (Washington, DC: USAID, 2020), [https://pdf.usaid.gov/pdf\\_docs/PA00X6M7.pdf](https://pdf.usaid.gov/pdf_docs/PA00X6M7.pdf).
- 83 RTI International, *Measurement and Research Support to Education Strategy Goal 1* (Washington, DC: USAID, 2016), <https://shared.rti.org/content/development-and-pilot-testing-additional-subtasks-early-grade-reading-assessment-egra-20>.
- 84 Uwezo Uganda, *Are Our Children Learning? Annual Learning Assessment Project* (Kampala: Uwezo Uganda, 2014).
- 85 Aser Centre, "Reading Assessment Module," [http://img.asercentre.org/docs/Aser%20survey/Tools\\_Testing/Reading/english.jpg](http://img.asercentre.org/docs/Aser%20survey/Tools_Testing/Reading/english.jpg).
- 86 Molly Jamieson Eberhardt, *Bringing Learning to Light: The Role of Citizen-led Assessments in Shifting the Education Agenda* (Washington, DC: Results for Development Institute, 2015).
- 87 Brian Wilcox, *Making School Inspection Visits More Effective: The English Experience* (Paris: UNESCO, 2000).
- 88 CAMFED, "Our Data, Our Destiny: Young Women Deliver Community Accountability to Girls" (May 18, 2016), <https://camfed.org/latest-news/young-women-deliver-data-accountability/>  
<https://camfed.org/latest-news/young-women-deliver-data-accountability>.
- 89 RTI International, *USAID's Early Grade Reading Program (EGRP) in Nepal*, 2020.
- 90 Jennifer Bowers, interview with Soma Omenye staff (2021).
- 91 Melissa Chiappetta, interview with Jeremy Koch, Chief of Party for UNICEF/USAID Yesuni Unuzira Activity (January 2021).
- 92 Wilcox, "Making School Inspection Visits More Effective," 2000.
- 93 Henry Braun, Anil Kanjee, Eric Bettinger, and Michael Kremer, *Improving Education through Assessment, Innovation, and Evaluation* (Cambridge, MA: American Academy of Arts and Sciences, 2006).
- 94 LQAS is a sampling methodology that uses small sample sizes to classify local areas as meeting or not meeting minimum performance standards. See *Toolkit for the Local Education Monitoring Approach* for more information: <https://shared.rti.org/content/local-education-monitoring-approach-lemma-toolkit>.
- 95 Elks, "The Impact of Assessment Results," 2016; Raudonyte, "Use of Learning Assessment Data," 2019.
- 96 Evelyn Jekemei, "The Power of Instructional Support: Using Existing Systems to Change Teacher Behavior in Kenya," presentation for the USAID Tusome Early Grade Reading Activity.
- 97 Interview with Central Square Foundation, 2021.
- 98 Interview with Jeremy Kock, Abt Associates (2021).
- 99 Julius Atuhurra and Michelle Kaffenberger, "System (In)Coherence: Quantifying the Alignment of Primary Education Curriculum Standards, Examinations, and Instruction in Two East African Countries," RISE Working Paper 20/057 (December 2020).
- 100 Luis Crouch and Joe DeStefano, "A Practical Approach to In-Country Systems Research" (June 2015).
- 101 Management Systems International, *Policy Linking for Measuring Global Learning Outcomes Toolkit: Linking Assessments to the Global Proficiency Framework* (USAID, UNESCO, Global Alliance to Monitor Learning, 2020).
- 102 Anne Hayes, Ann Turnbull, and Norma Moran, *Universal Design for Learning to Help All Children Read: Promoting Literacy for Learners with Disabilities* (Washington, DC: USAID, 2018).
- 103 Luis Crouch and Joe DeStefano, "A Practical Approach to In-Country Systems Research" (June 2015).
- 104 Teaching at the Right Level, "TaRL in Action," <https://www.teachingattherightlevel.org/tarl-in-action>.
- 105 Gove, *Early Grade Reading Assessment (EGRA) Toolkit*, 2015.
- 106 Ibid.
- 107 María José Ramírez, *Quick Guide No. 3: Implementing a National Learning Assessment* (Quebec: UNESCO Institute for Statistics, 2017).
- 108 Braun et al., *Improving Education*, 2006.
- 109 Kellaghan, Greaney, and Murray, *Using the Results of a National Assessment of Educational Achievement*, 2009.
- 110 Anikita Bhatkande, "Schools in Maharashtra Asked to Do Self-Assessment under Centre's Initiative," *Hindustan Times* (November 24, 2019).
- 111 Kellaghan, Greaney, and Murray, *Using the Results of a National Assessment of Educational Achievement*, 2009.
- 112 RTI International, *USAID's Early Grade Reading and Mathematics Initiative (RAMP) in Jordan: Year 3 Progress Report; October 2016–September 2017* (Washington, DC: USAID, 2017), [https://pdf.usaid.gov/pdf\\_docs/PA00T7P.pdf](https://pdf.usaid.gov/pdf_docs/PA00T7P.pdf).
- 113 Yvonne Cao, "Interactive Dashboards for Adaptive Management: Enabling Differentiated School Support across 100 School Districts in Ghana" (2019).
- 114 Kellaghan, Greaney, and Murray, *Using the Results of a National Assessment of Educational Achievement*, 2009; T. S. Powdhyel, "The Bhutanese Education Assessment Experience: Some Reflections," *Prospects* 35, no. 1 (2005): 45–57
- 115 Eberhardt, *Bringing Learning to Light*, 2015; RTI International, *USAID's Early Grade Reading Program (EGRP) in Nepal*, 2020.

