# POLICY LINKING OVERVIEW

## OVERVIEW:

Policy linking is a standard-setting methodology, long used in many countries, to set benchmarks (or cut scores) on learning assessments that allow those assessments to be aligned across countries and contexts. While it is an old standard-setting methodology, its use has been extended to help countries set benchmarks that will allow reporting against global standards. Policy linking allows countries to use their existing national assessments or early grade reading and math assessments to report against the USAID Foreign Assistance ("F") indicators and the Sustainable Development Goals (SDGs). It works by linking national assessments to the Global Proficiency Framework (GPF), a framework developed by multilateral donors and partners based on current national content and assessment frameworks across more than 100 countries. The GPFprovides performance expectations/ standards for learners in Grades 2-6 in reading and mathematics. By linking their national assessments to the GPF, countries and donors are able to compare learning outcomes across language groups in countries as well as across countries and over time, assuming all new assessments are subsequently linked to the GPF.

## CHALLENGES WITH COMPARING/AGGREGATING LEARNING OUTCOMES:

Countries face three common challenges to ensuring both the comparability of reading and mathematics outcomes and their ability to aggregate outcomes from both national and other representative assessments to inform programming:

1) **Language of Assessment -** First, they have long faced a challenge in understanding how learners who are learning in different languages are performing in relation to one another and, thus, in determining where to invest scarce resources to have the most impact to improve learning. This is because the rate of acquisition of these languages and the ability to learn to read in these languages differs across languages. It takes less time for learners to learn to read in languages with more transparent orthographies (languages with fewer sounds and direct correspondence between sounds and letters), such as Spanish, than it does for learners to learn to read in more complex languages, such as Arabic or Amharic.
2) **Changes in Instruments Over Time -** Second, many countries have also faced challenges in comparing assessment results over time due to changing assessment instruments, which affects the difficulty level of the assessments, and not having equated the different tests.
3) **Comparing with Other Countries -** Finally, countries have struggled to compare their learning assessment results with results from other countries to gain an understanding of how their systems and learners stack up against those of other countries given the challenges described above with both differing languages of assessments and different difficulty levels of assessment instruments.

Similarly, international donors face significant challenges in identifying in which countries the greatest need for education interventions lies, which groups are most in need of targeted interventions within countries, and how much progress is being made comparatively across contexts and populations toward improving children's reading and math abilities. This lack of comparability also makes country reporting against SDG 4.1.1 infeasible.

## SOLUTIONS:

The best way to address these challenges is linking national assessments across languages and contexts. There are two ways of linking assessments with different characteristics:

1) **Statistical linking –** Statistical linking can take two forms—test-based linking and item-based linking. With test-based linking, the same learners take two different assessments, which can then be calibrated to ensure equivalent levels of difficulty. With item-based linking, a subset of items from multiple assessments are included in the same assessment to allow for calibration across assessments. This is more suitable to new data collection where linking is envisioned as part of the design and requires more resources and political buy-in.

2) **Policy linking –** Policy linking is a form of linking in which experts judge the level of difficulty of assessment items using a common set of performance standards (like the GPF) and set corresponding benchmarks to define levels of minimum proficiency. Assessment results can be compared across assessments at the benchmarks for each assessment that has been linked to the standards (or GPF).

Statistical linking tends to be costly and time consuming, as it requires learners taking multiple tests or longer tests across large samples. It also requires engagement of psychometricians with experience statistically calibrating assessments. Given these constraints, policy linking has gained international buy-in as a cost-effective and practical solution for countries looking to maintain use of their national assessments to set benchmarks that will allow them to report against USAID's F Indicators and the SDGs and compare to results over time, across contexts, and across countries.

### IMPLEMENTING POLICY LINKING:

Policy linking works by bringing together a group of approximately 15 panelists per assessment that includes expert teachers, content specialists, and language experts from across a country to collectively set benchmarks. The panel reviews the GPF, which defines performance standards for learners under four global proficiency levels (GPLs)—does not meet minimum proficiency, partially meets minimum proficiency, meets minimum proficiency, and exceeds minimum proficiency—with three benchmarks, as shown in the figure below. Using the detailed descriptors in the GPF, the panelists determine alignment between the GPF and the items that appear in their national assessment. Assuming strong alignment (which means that the majority of subconstructs are covered at least partially by the assessment items), the panelists continue forward to the next step of policy linking. In the second step, the panelists each independently determine how learners who fall into each GPL would likely perform on each item from the national assessment being linked. They then review how panelist scores compare and are presented with impact data that how the percentage of learners who would fall into each GPL based on their ratings. Finally, the panelists independently review each assessment item again and determine whether learners in each proficiency category would get the item correct or incorrect. The facilitators then compile and average the benchmarks identified by the panelists and present them to the Ministry of Education for reporting against global indicators.
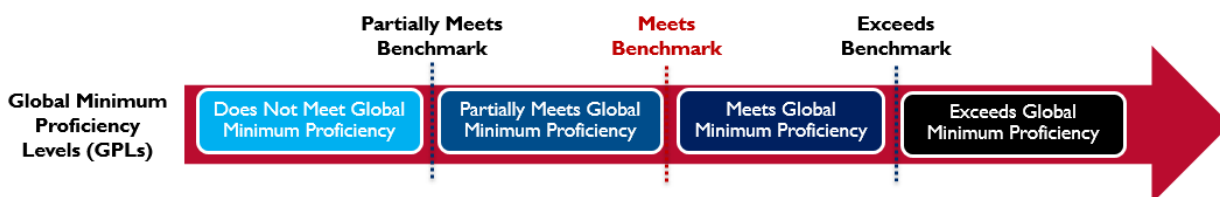


### REQUIREMENTS AND COSTS:

- Government buy-in and commitment to conducting the workshop and using the benchmarks
- 12 master teachers and 3 curriculum/teacher training/pedagogy specialists per grade level/language/subject combination for which officials wish to set internationally linked benchmarks (note: 2-3 grade-level/language/subject benchmarks can be set during a 4-day workshop, e.g., in a four-day workshop, a country could set Grade 2 reading and mathematics benchmarks for assessments administered in one language); these teachers and experts will act as the workshop panelists
- Expert facilitators (a lead facilitator and one content facilitator with expertise in reading or math and strong familiarity with the GPF per panel, along with a data analyst)
- A recent quality reading and/or mathematics assessment that aligns closely with the GPF

- Approximately $100,000 per workshop to facilitate panelist transportation and per diem costs as well as cover facilities and snacks

## ADVANTAGES:

- Low-cost method for countries to set benchmarks that are linked to global standards
- Allows countries to continue to use their national assessments for measuring learning outcomes and reporting on global indicators
- Supports wide government buy-in and ownership since benchmarks are set by countries rather than donors and can be continued by governments after development assistance ends
- Does not rely on current learner performance to set benchmarks, which can distort benchmarks when there are not a sufficient number of proficient learners in a country
- Allows countries to compare assessment results over time, across contexts, and across languages to better target interventions to the most vulnerable groups (when all assessments are linked to the GPF)
- Allows countries to compare progress in learning outcomes amongst one another
- Allows countries to report against USAID's "F" indicators and the SDGs

## NATIONAL VERSUS INTERNATIONAL BENCHMARKS:

- Countries may choose to set national benchmarks that vary from their internationally linked ones. They may wish to do this if they have an established content and/or assessment framework that does not align with the GPF or if they do not agree that the GPF standards are appropriate in their context.
- If the country has not already set national benchmarks, they could use the same policy linking workshop to set both national and international benchmarks, using their national content and/or assessment framework and the GPF in the same workshop to generate two sets of benchmarks.

## KEY TERMS:

**Content Standards** – describe what learners should know and do as a result of instruction.  Countries typically set their own content standards.  USAID and other stakeholders used content standards from more than 100 countries to inform the GPF.

**Performance Standards** – describe how learners should perform on assessments to meet the content standard; the GPF includes the internationally-agreed-upon performance standards for grades 2-6 in reading and math.  Countries can also set their own performance standards and set benchmarks for those standards; they just will not be able to use those standards to report against the SDGs.

**Benchmark** - a specific assessment score that designates a performance standard has been met on a given assessment; the desired competence or skills learners should developmentally be able to demonstrate when provided appropriate resources and support. Benchmarks should be set based on objective evidence of required performance rather than the current performance of learners.

**Target** – a goal for the number or percentage of children that will reach the benchmark for a given grade in a given period of time; targets should be altered based on the current performance of learners in schools at that grade and should provide a realistic timeline for when learners at that grade should be able to achieve minimum proficiency standards. They should also vary across populations.

**Global Proficiency Framework (GPF) –** a framework developed by donors and partners based on current country content and assessment frameworks across more than 100 countries, that provides performance expectations/standards for learners in Grades 2-6 in reading and mathematics

**Global Proficiency Levels (GPLs) –** the four levels of proficiency defined by the GPF for classifying learner outcomes: does not meet minimum proficiency, partially meets minimum proficiency, meets minimum proficiency, and exceeds minimum proficiency.