The challenge: assessing learning outcomes across different contexts

Assessing learning outcomes across nine countries

As the Evaluation Manager of the UK Department for International Development’s (DFID’s) £355 million Girls’ Education Challenge Fund (GEC), Coffey is measuring the fund’s impact on the learning outcomes of children across 17 countries. Projects funded through GEC’s Step Change Window aim to quickly and effectively expand learning opportunities for 650,000 girls at primary and secondary school levels in nine focus countries.

One of Coffey’s requirements for the Step Change Window evaluation is to assess literacy and numeracy outcomes through learning assessment tools in a way that is consistent and comparable across the funded projects. A key challenge is that project interventions are taking place in a wide range of contexts involving a variety of target populations whose specific social and institutional contexts need to be accounted for as they may significantly influence girls’ learning outcomes. In addition, existing local learning assessments may not be reliable guides to measure progress, and thus the impact of the projects.

“A key challenge is that project interventions are taking place in a wide range of contexts involving a variety of target populations whose specific social and institutional contexts need to be accounted for as they may significantly influence girls’ learning outcomes.”

To take up this challenge, Coffey has developed and implemented an ambitious program of research activities, including an independent assessment of literacy and numeracy outcomes. We found a way to harmonise learning outcomes across contexts while ensuring that they reflected as accurately as possible the progress of girls compared to other girls tested within their context.

Complementing projects’ research with independent rigorous program-wide research

The GEC evaluation strategy required the funded projects to assess the literacy and numeracy outcomes of a cohort of girls. The projects could choose from a range of international learning assessment tools. Most projects used the Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA) tools. Some used the Annual Status of Education Report (ASER) or Uwezo, or tests derived from national examinations. The variety of these

---

1 EGRA and EGMA have been developed by our partner RTI International: [http://www.rti.org/](http://www.rti.org/). EGRA is made of four subtasks: letter recognition, phonemic awareness, oral reading and reading comprehension. EGMA includes number recognition, comparisons, addition, subtraction and additional calculation subtasks.

2 ASER was developed in India to test children aged 6-16 years. Children are assigned a competency level between 1 and 5 for literacy (from reading word to reading story) and between 1 and 7 for numeracy (from number recognition to division). They are marked at the highest level which they can perform comfortably. Uwezo means ‘capability’ in Kiswahili and is based on an ASER tool and was originally developed for use in Kenya.
assessments was a major constraint to the comparison of learning outcomes across contexts.

Coffey therefore led extensive baseline research to supplement the projects’ data collection activities. In each of the 15 project areas, local enumerators conducted about 400 structured household interviews within which a randomly selected girl was assessed using the EGRA and EGMA tools. The local interviewers used a consistent template across all of the surveyed areas, which had been translated into the relevant local languages (usually the language of instruction).

**Developing a learning score that is comparable across contexts**

The administration of the EGRA tool in different languages places a limit on the comparability of literacy scores across contexts. It is possible to report scores from just one of the EGRA subtasks, reading fluency, to allow comparisons with international norms expressed in terms of words read per minute. However doing this ignores other aspects of literacy ability captured by subtasks other than words read per minute, such as letter recognition or reading comprehension. Taking the other subtasks into account can help support a fairer comparison between contexts, because the rate of progress on specific subtasks may vary by language.

As a consequence, Coffey has implemented an innovative statistical-based approach to develop a single EGRA score that enables better and fairer comparisons of reading fluency across projects in the various countries. Our model enables us to evaluate the relative difficulty of each subtask based on girls’ EGRA scores within each context. The relative difficulties of these subtasks are then used to weight and aggregate subtask scores into a single literacy indicator. Finally, the scale of the indicator is adjusted to the reading fluency score so that it is comparable with international benchmarks.

As it draws on all elements of the EGRA test, our aggregated literacy score also allows us to better discriminate at the bottom end of the spectrum. For instance, two students who are unable to read a single word have the same reading fluency score of zero word-per-minute. Provided they score differently in other subtasks, our model will give them different aggregated scores. This assessment of learning based on a wider range of ability will also allow us to more precisely evaluate the progress of low-performing students across time, especially as the GEC aims to help those who are most educationally deprived.

---

Tanzania and Uganda to assess whether children can perform literacy and numeracy skills at a primary grade 2 level of difficulty.
3 Research on early development of reading skills suggests that all children move through the same stages when learning how to read. The pace at which they move through these stages may differ by language, which is another challenge that we have had to deal with.
4 Our approach is based on the item-response theory which associates a person’s response to a multiple item test with a single ability – or trait – score.
How we can help

Coffey has invested strategic thinking into getting more reliability and value from the evidence that we generate, and in maximising the usefulness of the data for the specific populations that projects are most seeking to help.

Coffey's expertise in assessing learning outcomes for the GEC can be useful to any program or project that aims to have a measurable impact on the education of different populations in different national or local contexts.

If you have any questions about how this might help you with your project, please feel free to send us an email at GECanalysis@coffey.com.