

SDG 4.1.1 Quality Assurance Board Terms of Reference

April 17, 2020

Overview: The United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute of Statistics (UIS) is seeking fifteen representatives to sit on the newly created Sustainable Development Goal (SDG) 4.1.1 Quality Assurance Board, which will be tasked with reviewing country assessments, sampling methodologies, linking methodologies, and other assessment and methodological materials to determine if country reporting meets UIS' standards and make recommendations to UIS about whether they should accept the reported results. Additionally, they will make recommendations back to countries that don't meet the criteria on what changes they might need to make to ensure their results can be accepted and also how they might improve their overall outcomes. The main indicator used to measure SDG 4.1.1 is *proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex.*

Background: UNESCO, the World Bank, the United States Agency for International Development (USAID), United Kingdom's Department for International Development (DFID), the Bill and Melinda Gates Foundation, and other international donors face significant challenges in identifying in which countries the greatest need for education interventions lies and how much progress is being made comparatively across countries in improving children's reading and math abilities. The reason for the challenge is that countries largely use different reading and math assessment instruments to monitor learning trends in their education systems. In addition, even when countries use the same assessments, the language of the assessment may still differ. This makes both fair and accurate comparisons between country outcomes and country reporting against SDG 4.1.1, as well as similar USAID and DFID global education indicators, extremely difficult. UIS has determined that the most feasible way to address this challenge is by linking national assessments.

There are two main types of linking: statistical and non-statistical (based on the judgements of experts). Statistical linking is generally more accurate, but it has greater requirements and requires a level of technical expertise that several countries lack, e.g., either common students taking two or more national assessments or having a fair number of common items embedded across two or more assessments in different countries. As with any process that involves the judgement of experts, non-statistical linking (called policy linking from here forward) is vulnerable to subjectivity. However, experts have agreed that it is acceptable when requirements for statistical linking are not met due to issues such as design, logistics, or cost. The adoption of policy linking as a method for linking results to SDG 4.1.1 helped UIS to upgrade the indicator from Tier 3 to Tier 2 and eventually Tier 1,¹ meaning that there is an established methodology for measuring against the indicator, and more than 50 percent of countries say they will be able to report outcomes.

In brief, policy linking is a method in which the link from one assessment to another is the same set of descriptions of levels of learner proficiency. Linking two assessments using the same set of competence levels and descriptions results in comparability through equivalent but different passing scores for assessments, e.g., a passing score of 55 on one (easier) assessment that is equal to a passing score of 40

¹ Tier 1: Indicator is conceptually clear, has an internationally established methodology and standards are available, and data are regularly produced by countries for at least 50 per cent of countries and of the population in every region where the indicator is relevant; Tier 2: Indicator is conceptually clear, has an internationally established methodology and standards are available, but data are not regularly produced by countries; and Tier 3: No internationally established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested.

on another (more difficult) assessment. Experts created the Global Proficiency Framework (GPF) in the spring of 2019 to establish global levels and descriptions of expected student performance to facilitate this linking. The GPF includes four performance levels, called the global proficiency levels (GPLs)—does not meet minimum proficiency, partially meets minimum proficiency, meets minimum proficiency, and exceeds minimum proficiency and then detailed proficiency descriptors (called the global proficiency descriptors or GPDs) for grades 2-6 in reading and math by domain, construct, and sub-construct.

The GPF has also become the framework to which countries are statistically linking assessments. And, UIS is using it as the framework for classifying assessment items it is curating to populate a Global Item Bank that countries will be able to use to generate new assessments that align with SDG 4.1.1 or to augment their existing assessments to improve alignment with the GPF and global standards.

Purpose of this Board: The purpose of the board is to review the rigor and quality of statistical linking and policy linking results reported to UIS for SDG 4.1.1 to ensure the validity and comparability of outcomes and to make recommendations to UIS on whether to accept individual reports for use in reporting against SDG 4.1.1. Additionally, they will make recommendations back to countries that don't meet the criteria on what changes they might need to make to ensure their results can be accepted and also how they might improve their overall outcomes.

Responsibilities:

- Task 1 - The board will review the draft quality assurance policy for policy linking and provide feedback and recommendations to be incorporated into the policy and pilot tested.
- Task 2 - The board will provide guidance and feedback on the generation of quality assurance policies for statistical linking and item bank approaches.
- Task 3 – The board will provide quality assurance for the classification of items in the item bank.
- Task 4 - Collectively, the board will review all country proposals for linking assessment results to SDG 4.1.1 pre-implementation of the approach to provide recommendations on how they might improve the methodology and make a recommendation to UIS on whether the country's assessment(s) meet(s) quality criteria
- Task 5 - They will review post-implementation reports that include details on processes and outcomes to make a recommendation to UIS on whether the results should be accepted for reporting against SDG 4.1.1.
- Task 6 – The board will make recommendations back to countries that don't meet the criteria either pre- or post-implementation to tell them what changes they might need to make to ensure their results can be accepted and also how they might improve their overall outcomes.

See Annex A for a draft diagram of the process for policy linking. The board will work with UIS and its partners to develop similar approaches for statistical linking and use of the item bank. Five board members will review each proposal and will independently grade the proposal and results against a detailed set of criteria on a four-point scale.

For policy linking (which is the method for which UNESCO is furthest along in establishing quality assurance guidelines), the board members will review country-submitted data and evidence at two points in time: 1) before a country engages in a policy linking workshop and 2) following the workshop to ensure the quality of the results.

Before a country engages in a policy linking workshop, board members will be expected to review:

- Alignment between the assessment, the assessment framework, and the curriculum

- Appropriateness of the assessment for the population and representativeness of the sample (if relevant)
- Reliability of the assessment

Following the workshop to ensure the quality of the results, board members will be expected to review the reliability and validity of the linking methodology and include examining the profile and representativeness of the panelists involved as well as outputs such as inter- and intra-rater reliability, standard error of measurement, and panelist confidence in outcomes.

The draft policy linking quality assurance policy provides criteria for scoring these materials using the four-point scale below (the board will need to draft similar methods and scales for the other methodologies, or a scale that can be applied to all three methods will need to be adopted):

- **(4) Excellent** - all criteria met with suitable responses for all questions
- **(3) Good** - all criteria met with suitable responses for the most critical questions, including adequate inter-rater reliability scores during the administration/ scoring of the assessment
- **(2) Sufficient** - all criteria met with suitable responses for all critical questions but adequate inter-rater reliability scores only calculated during enumerator or rater training
- **(1) Insufficient** - insufficient criteria met to continue with proposed method or for results to be accepted

Board members will be able to score proposals from wherever they are located (home or office). Scores will be averaged to determine the final rating, and all proposals and results with a score of 2 or above will be accepted. Detailed requirements for each criterion are included in the policy linking quality assurance policy, which will be shared following initial expressions of interest.

In addition to reviewing proposals and results, members will be required to participate in four-six board meetings per year. Most of these will take place remotely, but one or two may be in person.

Qualifications: Board members must meet one of the following sets of qualifications:

Psychometricians:

- Trained psychometrician
- Experience designing and administering assessments in developing countries
- Experience analyzing assessment results and access to software that allows for such analysis
- Experience conducting standard-setting benchmarking workshops (preferably in developing countries, but experience in any/all countries will be considered)
- Experience working in developing countries
- Strong communication skills

Content Specialists:

- Reading or math expert
- Experience teaching students in grades 2/3, end of primary, and/or early secondary school
- Experience setting content and/or performance standards or designing curricula
- Experience working in developing countries
- Strong communication skills
- Experience teaching, working in non-Latin-based languages, preferred (at least some board members must have this experience, including experience with non-Alphabetic languages)

Level of Effort: Time (level of effort) requirements for participants are approximately 10 days per quarter/40 days per year with the bulk of this effort required in March, June, September, and December.

Compensation: Board members will be compensated through a quarterly honorarium. Further, UIS will cover all travel and per diems costs associated with in-person board meetings.

Expression of Interest: Interested individuals are invited to submit their CVs and a short Letter of Interest to UIS at UIS.io@unesco.org by **Friday, June 19**. Applicants should demonstrate an understanding of the SOW, capability, availability, interest, and value. UIS will then follow up with qualified individuals.

Annex A – Overview of policy linking quality assurance process

