



United Nations  
Educational, Scientific and  
Cultural Organization



UNESCO  
INSTITUTE  
FOR  
STATISTICS



TECHNICAL  
COOPERATION  
GROUP



# SDG indicator 4.5.2:

## Recommended approaches for measurement

TCG Fifth Meeting  
15-16 November 2018  
Mexico City, Mexico  
**TCG5/REF/5**

## 1. Introduction

This note presents recommendations for the thematic SDG indicator 4.5.2: percentage of students in primary education whose first or home language is the language of instruction. The document proposes generic definitions and a generic formula, and offers recommendations with respect to the international implementation of indicator 4.5.2 with existing data sources.

## 2. Monitoring the indicative strategies to foster bi- and multilingual education

According to the Incheon Declaration, language policies in education are one of several strategies to address exclusion.<sup>1</sup> "... teaching and learning in the first or home language should be encouraged". This leads to the indicative strategy **"to foster bi- and multilingual education, starting with early learning in the first or home language of children"**.<sup>2</sup>

This is in accordance with UNESCO Resolution 30C/12, which recommends promotion of access to knowledge independent of the mother tongue. Bi-lingual education should begin in pre-primary and ideally involve the **use of two languages at primary-school level as "a medium of instruction, thus using two languages for the acquisition of knowledge"**.<sup>3</sup>

Thematic indicator 4.5.2 seeks to chart progress in language policies. It is obvious that progress to provide education for more children in their home language would transform education systems further to multilingual<sup>4</sup> systems with several languages of instruction. Thus, it is crucial that indicator 4.5.2 is applicable to multilingual systems.

**Recommendation 1:** The generic definitions underlying indicator 4.5.2 should be in accordance with the statistical requirements of multi-lingual education systems.

## 3. Definitions

### Language of the children: Home or first language(s)

The proposed label for the indicator 4.5.2 refers to children's "first or home language". The UNESCO-UIS<sup>5</sup> currently defines "First or home language as the student's main language of communication outside the school environment. It is usually either the first language students learned or the language of their family or local community." The definition accommodates several similar but not identical concepts of children's languages: mother tongue, language learned in early years, language spoken at home, language spoken with parents, local/community language. Most commonly used in existing

---

<sup>1</sup> UNESCO, 2015, p. 3.

<sup>2</sup> UNESCO, 2015, p. 38.

<sup>3</sup> UNESCO, p. 35.

<sup>4</sup> From this point on the term multilingual is intended to include bilingual.

<sup>5</sup> UNESCO Institute for Statistics, 2017a.

surveys is "home language" followed by "first language".<sup>6</sup> Results on 4.5.2 depend on the chosen concept and exact wording of the items used.<sup>7</sup> None of the potential data sources analysed refer to the language of the local community.

**Recommendation 2:** Narrow the definition of children's language to commonly used concepts in existing surveys: No reference should be made to the language of the local community. This also avoids ambiguities, when different languages are spoken at home and on the streets.

**Recommendation 3:** Harmonize between data sources: Different terminology should not be portrayed as interchangeable but as options. One concept should be marked as preferred: e.g. "current language use of children at home" and not "mother tongue or first learned language". "Use of language at home" should be given preference over questions on "language use in everyday life or out of school".

Shortening the definition of 4.5.2 to use "home language" as the main term could be considered.

Children in multilingual environments and migrant families use more than one language in their homes, and only few report to never speak the language of instruction at home. They are typically multilingual. The definition of home language should make some reference to the frequency of language use.

**Recommendation 4:** Define "home language" as the language spoken most often at home. Surveys recording the frequency of use of the language of instruction at home should provide the options "always" and "almost always", in addition to "sometimes" and "never".<sup>8</sup> The use of these four categories is recommended.

## Language(s) of instruction

The language of instruction refers to the medium used to teach the basic curriculum of the educational system. A language of instruction is one that is used for teaching *through* this language and not just teaching of this language.<sup>9</sup> Thus, "languages of instruction to teach content" should be distinguished from "languages taught as subject". The latter includes also reading/writing mother tongue as special subject, which is widespread.<sup>10</sup>

**Recommendation 5:** "Language of instruction" refers to languages used for teaching the basic curriculum of the educational system. Languages used only in language subjects, including mother tongue reading/writing are not languages of instruction, when not used to teach other content/subjects.

Recommendation 5 would imply that efforts of countries to teach reading and writing in mother tongues would remain invisible in 4.5.2, as long as no switch is made to teach *through* those languages.

<sup>6</sup> UNESCO Institute for Statistics, 2017b.

<sup>7</sup> UNESCO Institute for Statistics (2017b) provides a list of example questions and discusses measurement issues.

<sup>8</sup> Using too few options (e.g. only three as in PIRLS 2006) turns out problematic. See UNESCO Institute for Statistics (2017b, p. 20).

<sup>9</sup> UNESCO, 2003.

<sup>10</sup> UNESCO Institute for Statistics, 2017b, p. 12.

## Language(s) of instruction in bi- or multilingual settings

UNESCO support for multilingual education<sup>11</sup> should be reflected in the definition of indicator 4.5.2. The current wording pointing to “the” language of instruction implies one single language of instruction. Yet, four situations can occur:

- 1) *The child has a home language that is used as language of instruction:*
  - a) It is the single language of instruction.
  - b) It is one of several languages of instruction.
- 2) *The language(s) of instruction are not the child's home language(s):*
  - a) The home language is offered as subject (reading/writing).
  - b) The child's home language is not used in school.

**Recommendation 6:** To capture the language of instruction in multilingual systems, the above-mentioned situations 1.a) and 1.b) both should be counted as “providing education in the home language”. The definition of 4.5.2 should be clarified:

***“the percentage of students in primary education who have their first or home language as language of instruction”***

This allows for more than one language of instruction and focuses on the match of the home language with at least one of the languages of instruction.

In the metadata for indicator 4.5.2, “language of instruction” should be replaced by “language(s) of instruction”.

## Transitional multilingual education

The target population for indicator 4.5.2 are all children in primary education. Yet, reporting should be aware of transitional multilingual education with a planned transition from one language of instruction to another. Assume a system that offers 100% of students instruction in their home language in the first half of all grades of primary education, and that offers 100% of students instruction in another language in the second half of all grades of primary education. Such a system would for a given reference year be correctly reported as providing instruction in the home language to just 50% of total primary students. Yet, the fact, that the system ensures initial instruction in the home language for all students would be hidden.

**Recommendation 7:** The target population for the indicator should be the complete primary level. Reporting of 4.5.2 should be done for the complete level and separately for early grades (ideally grades 1 to 3) and grades towards the end of primary education to better reflect the situation in systems with transitional multilingual education.

<sup>11</sup> UNESCO (2003)

Data show that the use of home languages as medium of instruction is more frequent in lower grades than in higher grades. Using data from studies of only lower or upper primary grades would overestimate or underestimate indicator 4.5.2.

**Recommendation 8:** If data for a country exist only for a single grade of primary education, this grade should be clearly identified and estimations for the whole primary level should be avoided.

The interpretation of results for indicator 4.5.2 depends on understanding of the language policy in an education system. The appropriateness of data sources can only be judged when language policies are known.

**Recommendation 9:** Qualitative metadata should be collected on language policies identifying countries with different use of language of instruction in lower and higher grades.

## 4. Indicator definition

Indicator 4.5.2 should be calculated as follows.

**Label for thematic indicator 4.5.2:** The percentage of students in primary education who have their home language (or first language) as a language of instruction.

### Calculation method:

$$4.5.2 = PELA_{1,g} = EF_{1,g}/E_{1,g}$$

where:

**PELA<sub>1,g</sub>** = percentage of students in primary education (ISCED level 1), grade range *g* who have their home language (or first language) as a language of instruction

**EF<sub>1,g</sub>** = number of pupils in primary education (ISCED level 1), grade range *g* instructed using a home language as medium of instruction

**E<sub>1,g</sub>** = total pupils in primary education (ISCED level 1), grade range *g*

**Recommendation 10:** Data for indicator 4.5.2 should be labelled according to the above proposal. This ensures consistency with the above recommended definitions.

## 5. Implementation with existing international data sources

A number of international data sources can contribute to estimation of indicator 4.5.2. To ensure good coverage, data from multiple sources will need to be joined. An overview of sources is provided in UIS (2017b). Two types of sources can be distinguished: i) primary school surveys<sup>12</sup> and ii) household surveys.<sup>13</sup> This section will continue with specific recommendations on the use of different data sources.

<sup>12</sup> To be considered: PIRLS, TIMSS, PrePIRLS, SACMEQ, PASEC, LLECE, EGRA.

<sup>13</sup> To be considered: MICS, DHS, LSMS.

## Language of test

School surveys record the language of instruction not from children but generalized from process data. Mostly they do not record “language of instruction” but “language of test”. This can be problematic if languages of instruction are different for different subjects, i.e. in multilingual education.

**Recommendation 11:** When “language of test” is reported instead of “language(s) of instruction”, the data must be interpreted carefully when potentially more than one language of instruction is used in given schools. Data on multilingual education may have to be excluded if a student’s home language is used as language of instruction in other subjects but not in the test. Thus, to generalize from data on “language of test” to indicator 4.5.2 on “language of instruction” more knowledge of the system is required.

**Recommendation 12 (future development):** Assessment surveys should add questions on language(s) of instruction as an optional question to be used in multi-language systems. The LSMS could serve as an example for implementation of recommendation 13.<sup>14</sup>

## Coverage of total primary education

School surveys are samples of students of a single grade (sometimes two grades). TIMSS/PIRLS and SACMEQ are representative for higher primary grades, EGRA for lower primary grades.<sup>15</sup> PASEC and LLECE sample in lower and upper primary grades. As mentioned above, such data cannot be generalized to represent all primary-level students in systems with transitional multilingual education policies.

**Recommendation 13:** For systems without changes in the language of instruction from lower to upper grades, estimates for total primary education based on data for a single grade might be acceptable. For systems with a transitional language policy, this must be avoided. If samples of grades of lower and upper primary grades are available for a given country, the samples must be used for separate estimates of lower and upper primary grades, which then can be pooled.

Sample design is important to properly monitor equity. UNESCO<sup>16</sup> reports that survey designs often exclude especially vulnerable people. Children not speaking the language(s) of instruction are also more likely to be excluded from data sources used for 4.5.2. In TIMSS/PIRLS, schools with languages of instruction used in a small number of schools are excluded in many countries.

**Recommendation 14:** Sampling coverage of data sources needs to be reviewed to identify possible exclusions based on criteria related to language. If several data sources are available for a given country (e.g. household data and assessment data), the selection of the data source should be informed by coverage of language groups and schools.

<sup>14</sup> UNESCO Institute for Statistics, 2017b, p. 64.

<sup>15</sup> With exceptions.

<sup>16</sup> UNESCO, 2017, p. xvi.

## Questionnaire items

The exact questions used to capture home language (or similar concepts) can differ between survey types and between rounds of the same survey. This can have substantial effects on the capturing of children's home language or on their language of instruction.<sup>17</sup> Such uncertainty in data is a common problem and cannot be resolved when pooling different data sources.

**Recommendation 15 (development):** School assessment surveys should aim to reduce unnecessary variability in the way language of instruction and especially home language are captured. Two models could be chosen as examples:

- TIMSS-2015: based on a comparison of language of instruction and home language, without language list.
- EGRA/MICS: "Which language do you speak most of the time at home?" (With language list)

## Recommendations on pooling data sources to gain regional and global estimates

Regional and global estimates for indicator 4.5.2 will rely on a joined country-level database that is compiled from multiple data sources. The following steps are recommended:

**Recommendation 16:** For each potential international data source (e.g. TIMSS, MICS, SACMEQ, ...) all data points should be captured and documented. If available, national estimates should also be captured. The data should be stored and disseminated together with short remarks on coverage of the samples, definition of target populations, and grade levels represented.

Yet, situations must be avoided in which change over time in the indicator is related to changing data sources. Before pooling data for countries, for each country the one best data source should be identified and comparisons over time should keep the selected data source fixed.

**Recommendation 17:** Change over time in 4.5.2 should be monitored by using data from the same or at least strictly comparable survey(s).

Some criteria should be fixed to guide the selection of data sources if several data sources are available for a given country.

**Recommendation 18:** The following priorities should be considered when selecting data sources.

- National data should be used as first choice.
- Preference should be given to data sources covering all grades. Sources covering only lower or upper primary grades should only be used in the absence of sources with better grade coverage.
- The selection of the data source should be informed by sample design: coverage of language groups and schools.

<sup>17</sup> UNESCO Institute for Statistics (2017b, pp. 10-30): Questions can change over time, respondents can be children or parents or teachers, concepts can be similar but different.