UNESCO
United Nations
Educational, Scientific and
Cultural Organization

UNESCO
INSTITUTE
FOR
STATISTICS

SUSTAINABLE
DEVELOPMENT
GOALS

SDG 4 Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

METADATA

Target 4.5 By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Definition

Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction.
* Note that for the estimates derived from the student learning assessments, a proxy of language of instruction is used that is the language of the test.

Calculation method

For learning assessment data, the indicator is defined as the percent of students who speak the language of the test more than "sometimes" or "never", defined depending on the assessment (see Table 1). For assessment *i*, the measure of prevalence of learning in one's own language L_i in a particular country and sub-population would be defined as:

$$L_i = 100 \times E[l_i] \tag{1}$$

where l_i equals 1 if the student responded that he or she uses the language of the test more than "never" or "sometimes", 0 if he or she used the language of test "never" or "sometimes", and excluded if the student did not provide a valid answer. Table 1 presents the data sources, the questionnaire items used to develop the indicators, and the rules for determining whether a student is defined as learning in her or his home language or not.

Interpretation

For estimates using learning assessment data, the indicator provides the percent of students whose test language of test and language spoken at home are the same. This provides a

Updated March 2021

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

proxy to measure the percent of students learning in their home language as the language of the test is generally the language of instruction. However, it is not possible to verify (empirically) the actual language of instruction using learning assessments in this metadata as this data was not collected by these assessments. Earlier rounds of LLECE and SERCE did include a question on language of instruction; however, this assessment has not yet been added to the dataset used by UIS.

Disaggregation
By level of school, sex, urban or rural location and wealthiest and poorest 50 percent (see annex for details on definition of these sub-populations).

Metadata points: The metadata points indicate the source of data (Table 1 provides details for each data source). They also include standard errors and confidence intervals estimated based on the methodologies suggested by the assessment programme.

Data source: estimates provided by UNICEF based on MICS 6
Calculation method: Similar to equation (1), the estimation is based on the proportion of students reporting that the language used by teachers when teaching is equivalent to the most often language spoken at home. The population would be restricted to children currently attending primary school during the school year, as referred to in the MICS 6 questionnaire.

> **Measurement point definition:** The definition of measurement points follows those used by the Global Alliance to Monitor Learning (GAML) for learning assessment data (see Table 1 below):
> 1. Grade 2 or 3: Plus one year when primary lasts more than 4 years according to ISCED levels in the country
> 2. End of primary: Plus or minus one year from the last year of primary according to ISCED level mapping in the country
> 3. End of lower secondary: Plus two or minus one of last year of lower secondary according to ISCED level mapping in the country

Note that MICS 6 is assigned to end of primary. Table 2 summarize the definition in MICS.

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Table 1. Data sources and questions on use of test language at home

Data source	Target population	Language at home questions	Responses and mapping to whether the student uses the language of the test at home (yes/no/omitted)
LLECE 2013 (TERCE)	6th grade students (end of primary)	At home, which language do you speak most of the time?	"Spanish or Portuguese": yes All other valid responses: no Missing: omitted
PASEC 2014	2nd grade students (grades 2/3)	What language do you speak at home?	"You always speak <language of test>": yes "You speak <language of test> sometimes and another language sometimes": no "You never speak <language of test>": no Missing: omitted
PASEC 2014	6th grade students (end of primary)	What language do you speak at home?	"You always speak <language of test>": yes "You speak <language of test> sometimes and another language sometimes": no "You never speak <language of test>": no Missing: omitted

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Table 1. Data sources and questions on use of test language at home

Data source	Target population	Language at home questions	Responses and mapping to whether the student uses the language of the test at home (yes/no/omitted)
PISA 2018	15 year-old secondary students (end of lower secondary)	What language do you speak at home most of the time? (please select one response)	"<Language 1>" "<Language 2>" "<Language 3>" "<…etc.>" "Other languages" Assignment of these responses to whether the student speaks the language of the test at home most of the time is done by the OECD and reported as a variable in the dataset.
SEA-PLM 2019	5th grade students (end of primary)	What language do you speak at home most of the time? Note if two languages are spoken at the same frequency, choose the one you learnt first.	"<Language 1>" "<Language 2>" "<Language 3>" "<Language 4>" "<Other language>" Assignment of these responses to whether the student speaks the language of the test at home most of the time is done by SEA-PLM and reported as a variable in the dataset.

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Table 2. Data sources and questions on used in MICS 6 to estimate percent of children learning in their home language

DATA SOURCE	TARGET POPULATION	QUESTIONS ON LANGUAGE USED AT HOME AND LANGUAGE USED BY TEACHERS	RESPONSES AND MAPPING TO WHETHER THE STUDENT USES THE LANGUAGE OF THE TEST AT HOME (YES/NO/OMITTED)
MICS 6	Children aged 5 to 17	FL.7 Which language do you speak most of the time at home? FL9A. What language do your teachers use most of the time when teaching you in class?	Mapped to yes if language answered in both questions is equal; no if unequal, and omitted if either of the two questions has a missing or invalid response

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Annex: metadata for sub-population definitions from the cross-national learning assessments

Definition of sub-populations

Female and male: The dataset used to estimate the indicator includes a question asking whether the student is male or female. For TIMSS, the administrative record of the sex of the student was used following how TIMSS reports learning achievement scores by sex.

Urban and rural: All assessments ask the school director about the type of location in which the school is located; however, only LLECE 2013 asks explicitly whether the school is located in an urban or rural area. The other surveys ask the question in various ways included the number of inhabitants or by description. See Table A.1 for the questions from each assessment and how they were mapped to urban or rural.

High and low socioeconomic status: All assessments, except TIMSS, provide a measure of the socioeconomic status of students (SES). This is typically based on the responses from students about assets at home as well as the education of parents. LLECE 2013 used the responses of the family questionnaire to generate its index. PASEC 2014 and PISA 2018 used student responses; no index was generated for the PASEC 2014 2nd grade students given their young age and reliability of answers. TIMSS reports an index of home learning resources based on household possessions reported by students and it was used as a measure of socioeconomic status. The SEA-PLM generates the index based on parental education and home resources.

To define high and low SES students, the median was calculated for each country, student above the median were defined as high SES while those below were defined as low SES. See Table A.2 for the names of the variables used to define high and low SES in each assessment.

Non-response and small sample sizes: Indicator estimates were not reported for sub-populations if data for the sub-population was available for less than 90 percent of the sampled students or if the number of observations for a particular sub-population was less than 100.

Standard errors and confidence intervals methodology
The suggested methodology for estimating standard errors and subsequent confidence intervals varies by assessment and aims to account for clustering at the school-level. All surveys suggest using replicate methods in which the sample variation is obtained from

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

variously defined sub-samples that mimic the sample design; the variation in estimates among the replicates provides an estimate of the sampling variation. The suggested methods were used for all assessments except LLECE 2013. For this survey, replicate weights were provided with each of the learning achievement datasets; however, a large number of students in the background dataset (which included the responses to the bullying and home language questions) were not included in the student achievement dataset. In order to maximize the background data, a linearization method for estimating the standard errors robust to clustering at the school level was used. Table A.3 describes the methodology used for each assessment.

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Table A.1. Definition of urban and rural sub-populations

Assessment	Population	Question	Responses (mapping)
LLECE 2013	Grades 3 and 6	How would you characterize the area where your school is located?	In an area considered rural (**rural**) In an area considered urban (**urban**)
PASEC 2014	Grades 2 and 6	Your school is located in…	A town (**urban**) A suburb of a big city (**urban**) A big village (hundreds of homesteads) (**rural**) A small village (dozens of homesteads) (**rural**)
PISA 2018	15 year-olds	Which of the following definitions best describes the community in which your school is located?	A village, hamlet or rural area (fewer than 3 000 people) (**rural**) A small town (3 000 to about 15 000 people) (**rural**) A town (15 000 to about 100 000 people) (**urban**) A city (100 000 to about 1 000 000 people) (**urban**) A large city (with over 1 000 000 people) (**urban**)
SEA-PLM 2019	Grade 5	Which of the following characteristics best describes the community in which your school is located?	A village, or rural area (fewer than 3 000 people) (**rural**) A small town (3 000 to about 15 000 people) (**rural**) A town (15 000 to about 100 000 people) (**urban**) A city (100 000 to about 1 000 000 people) (**urban**) A large city (with over 1 000 000 people) (**urban**)
TIMSS 2015	Grades 4 and 8	Which best describes the immediate area in which your school is located?	Urban–Densely populated (**urban**) Suburban–On fringe or outskirts of urban area (**urban**) Medium size city or large town (**urban**) Small town or village (**rural**) Remote rural (**rural**)

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Table A.2. Variables used to define high and low socioeconomic status of students

Assessment	Population	Variable	Respondents
LLECE 2013	Grades 3 and 6	Index of the family's socioeconomic status (isecf)	Parents
PASEC 2014	Grade 2	n/a	n/a
PASEC 2014	Grade 6	Socioeconomic index of the student's family (ses)	Students
PISA 2018	15 year-olds	Index of economic, social and cultural status (escs)	Students
SEA-PLM 2019	Grade 5	Socioeconomic status index (ses)	Students and parents
TIMSS 2015	4th grade	Index of home resources for learning (asbghrl)	Students
TIMSS 2015	8th grade	Index of home educational resources (bsbgher)	Students

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

Table A.3. Methodology for calculating standard errors by assessment

Assessment	Method	Reference for formulas	Software routine
LLECE 2013	Linearized	StataCorp 2013	SVY module for Stata (StataCorp)
PASEC 2014	Jackknife repeated replication	PASEC 2017	PV module for Stata (Macdonald 2008)
PISA 2018	Balanced repeated replication	OECD 2009	PV module for Stata (Macdonald 2008)
SEA-PLM 2019	Jackknife repeated replication	SEA-PLM 2020	PV module for Stata (Macdonald 2008)
TIMSS 2015	Jackknife repeated replication	Foy & LaRoche (2016)	PV module for Stata (Macdonald 2008)

4.5.2 Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

References

Foy, P. and S. LaRoche (2016). Chapter 4. Estimating Standard Errors in the TIMSS 2015 Results. In. Michael O. Martin, Ina V.S. Mullis, and Martin Hooper, Editors, *Methods and Procedures in TIMSS 2015*. Boston College, TIMSS & PIRLS International Study Center available at: https://timssandpirls.bc.edu/publications/timss/2015-methods/chapter-4.html

OECD (2009). *PISA Data Analysis Manual Spss, Second Edition*. Paris: OECD. Available at: https://www.oecd-ilibrary.org/docserver/9789264056275-en.pdf?expires=1593772556&id=id&accname=guest&checksum=256201EAFD784D9168955CFDB6068789

PASEC (2017). *Manuel d'exploitation des données : Évaluation internationale PASEC2014*. PASEC, CONFEMEN, Dakar. Available at: https://www.pasec.confemen.org/wp-content/uploads/2017/06/Manuel_exploitation_donn%C3%A9es_%C3%A9valuation_internationale_PASEC2014.pdf

SEAPLM (2020). Data User Manual. Bangkok: SEAMEO and UNICEF

StataCorp (2013). *Stata Survey Data Reference Manual Release 13*. College Station, Texas, USA: Stata Press. Available at: https://www.stata.com/manuals13/svy.pdf