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

METADATA

Target 4.1 By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes

4.5.2 Percentage of students in primary education whose first or home language is the language of instruction

Definition
This indicator measures the percentage of students in primary education who have their first or home language as language of instruction.

Operational definition
First or home language is the language spoken most often at home.

Language of instruction is the language used for teaching the basic curriculum of the educational system.

Calculation method
For learning assessment data, the indicator would be 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 A.2 below). 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:

$$L_i = 100 \times E[l_i] \quad\quad (1)$$

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 she or he used the language of test never or sometimes, and excluded if the student did not provide a valid answer. For data based on MICS[1] 6 and subsequent rounds, the definition would be similar in which l_i equals 1 if the language most often used by teachers when teaching is equivalent to the language spoken at home most often. The population would be restricted to children currently attending primary school during the school year referred to in the MICS questionnaire.

---

[1] MICS: Multiple Indicator Cluster Surveys

Updated September 2020

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Interpretation

For estimates using learning assessment data, the indicator provides the percent of students whose test language and language spoken at home is the same. This provides a proxy for measure for 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 in the learning assessments in this meta-data as this data was not collected by these assessments. An earlier round of LLECE[2], SERCE[3], did include a question on language of instruction; however, this assessment has not yet been added to the dataset.

Type of data source

Data from household surveys collecting language most often spoken at home and spoken by teachers when teaching. Data from student assessments' background questionnaires on language spoken at home.

Disaggregation

By level of school, sex, urban or rural location and wealthiest and poorest 50 percent (see common metadata for indicators derived from student assessment data for details on definition of these sub-populations).

Data required

Table A1 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.

---

[2] LLECE: El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
[3] SERCE: Second Regional Comparative and Explanatory Study

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Data sources

Data on the administration of a large-scale assessment from a national representative sample from national learning assessment offices, ministries of education or other bodies responsible for learning assessments, including regional or international organizations running learning assessments (e.g. CONFEMEN[4], EQAP[5], IEA[6], OECD[7], SACMEQ[8] and TERCE[9]).

Metadata points

The metadata points indicate the source of data (Table A2 provides details for each data source). They also include standard errors and confidence intervals estimated based on the methodologies suggested by the assessment programme (see common metadata for indicators derived from student assessment data for details on methodology used to estimate standard errors and confidence intervals).

---

[4] CONFEMEN: Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie

[5] EQAP: Educational Quality & Assessment Programme

[6] IEA: International Association for the Evaluation of Educational Achievement

[7] OECD: Organisation for Economic Co-Operation and Development

[8] SACMEQ: Southern and Eastern Africa Consortium for Monitoring Education Quality

[9] TERCE: Third Regional Comparative and Explanatory Study

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Table A1. 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	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	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	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 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Table A1. Data sources and questions on use of test language at home

PISA 2018	15 year old secondary students	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.
TIMSS 2015	4th grade students	How often do you speak <language of test> at home? Fill one circle only.	"I always speak <language of test> at home": yes "I almost always speak <language of test> at home": yes I sometimes speak <language of test> and sometimes speak another language at home": no "I never speak <language of test> at home": no Missing or invalid: omitted
TIMSS 2015	8th grade students	How often do you speak <language of test> at home? Fill one circle only.	"Always": yes "Almost always": yes "Sometimes": no "Never": no Missing or invalid: omitted

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Table A2. 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 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Disaggregation
Definitions of sub-populations

Female and male: The dataset used to estimate the indicators include 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 about the type of location in which the school is located to the school director; 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 1 for the questions from each assessment and how they were mapped to urban or rural.

High and low SES: All assessments except TIMSS provide a measure of the socioeconomic status of students. This is typically based on the response by students about assets at home as well as 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 this was used as a measure of socioeconomic status. 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 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 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 aim to account for clustering at the school-level. All surveys suggest using replicate methods in which the sample variation is obtained from 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. To maximize the background data, a linearization method for estimating the standard errors robust to clustering at the school level was used. Table 3 describes the methodology used for each assessment.

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Table 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 old	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**)
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 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Table 2. Variables used to define high and low SES 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 old	Index of economic, social and cultural status (escs)	Students
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 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Table 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)
TIMSS 2015	Jackknife repeated replication	Foy & LaRoche (2016)	PV module for Stata (Macdonald 2008)

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Protocol for reporting

Some countries have more than one data source or estimates for different grades of primary education. The criteria used to select the source of data or grade to report on Indicator 4.5.2 prioritize the coverage of the survey (e.g. nationally representative surveys) and the earlier grades. The table below summarizes the decisions made with the data available for reporting.

Country	Coverage	Level of coverage	Target Grade	Year	Source	Value	Included in the data release?
Argentina	Argentina	National	Grade 6	2013	LLECE 2013 G6	98.38322	Yes
Argentina	Buenos Aires	Sub-national	Grade 4	2015	TIMSS 2015 G4	87.0518	No
Benin	Benin	National	Grade 2	2014	PASEC 2014 G2	1.358935	Yes
Benin	Benin	National	Grade 6	2014	PASEC 2014 G6	4.947303	No
Burkina Faso	Burkina Faso	National	Grade 2	2014	PASEC 2014 G2	0.4115603	Yes
Burkina Faso	Burkina Faso	National	Grade 6	2014	PASEC 2014 G6	5.832576	No
Burundi	Burundi	National	Grade 2	2014	PASEC 2014 G2	93.66841	Yes
Burundi	Burundi	National	Grade 6	2014	PASEC 2014 G6	3.049138	No
Cameroon	Cameroon	National	Grade 2	2014	PASEC 2014 G2	15.54811	Yes
Cameroon	Cameroon	National	Grade 6	2014	PASEC 2014 G6	15.70071	No
Canada	Canada	National	Grade 4	2015	TIMSS 2015 G4	75.26906	Yes
Canada	Ontario	Sub-national	Grade 4	2015	TIMSS 2015 G4	74.01808	No
Canada	Quebec	Sub-national	Grade 4	2015	TIMSS 2015 G4	77.62276	No
Chad	Chad	National	Grade 2	2014	PASEC 2014 G2	0.8990707	Yes
Chad	Chad	National	Grade 6	2014	PASEC 2014 G6	6.551772	No
Chile	Chile	National	Grade 6	2013	LLECE 2013 G6	98.63066	No
Chile	Chile	National	Grade 4	2015	TIMSS 2015 G4	89.67681	Yes
Congo	Congo	National	Grade 2	2014	PASEC 2014 G2	12.21658	Yes
Congo	Congo	National	Grade 6	2014	PASEC 2014 G6	20.81665	No
Côte d'Ivoire	Côte d'Ivoire	National	Grade 2	2014	PASEC 2014 G2	7.024707	Yes

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

Country	Coverage	Level of coverage	Target Grade	Year	Source	Value	Included in the data release?
Côte d'Ivoire	Côte d'Ivoire	National	Grade 6	2014	PASEC 2014 G6	12.50396	No
Mexico	Mexico	National	Grade 6	2013	LLECE 2013 G6	95.68874	Yes
Mexico	Nuevo Leon	Sub-national	Grade 6	2013	LLECE 2013 G6	99.28918	No
Niger	Niger	National	Grade 2	2014	PASEC 2014 G2	0.5114056	Yes
Niger	Niger	National	Grade 6	2014	PASEC 2014 G6	3.766104	No
Norway	Norway	National	Grade 4	2015	TIMSS 2015 G4	84.92798	Yes
Norway	Norway - 4th grade	Alternate grade	Grade 4	2015	TIMSS 2015 G4	82.69392	No
Senegal	Senegal	National	Grade 2	2014	PASEC 2014 G2	1.153053	Yes
Senegal	Senegal	National	Grade 6	2014	PASEC 2014 G6	5.795933	No
Togo	Togo	National	Grade 2	2014	PASEC 2014 G2	1.22698	Yes
Togo	Togo	National	Grade 6	2014	PASEC 2014 G6	6.339205	No
Togo	Togo	National	Primary	2017	MICS 6	3	No
United Arab Emirates	United Arab Emirates	National	Grade 4	2015	TIMSS 2015 G4	53.53459	Yes
United Arab Emirates	United Arab Emirates - Abu Dhabi	Sub-national	Grade 4	2015	TIMSS 2015 G4	56.40646	No
United Arab Emirates	United Arab Emirates - Dubai	Sub-national	Grade 4	2015	TIMSS 2015 G4	48.42261	No
United Kingdom of Great Britain and Northern Ireland	United Kingdom - England	Sub-national	Grade 4	2015	TIMSS 2015 G4	82.76352	Yes
United Kingdom of Great Britain and Northern Ireland	United Kingdom - Northern Ireland	Sub-national	Grade 4	2015	TIMSS 2015 G4	91.7889	No

4.5.2 Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education

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

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