

## SDG 4.5.2 Methodological Note

### SDG 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

August 2023

**Abstract:** This note presents methodology for estimating measures of SDG 4.5.2 using international student assessment data in addition to MICS 6. The primary limitation of using international assessment data is that language of instruction is not readily available in these datasets; rather, whether the language of the test is used at home is reported. Generally, the language of the test for international student assessments matches the official language of instruction. Comparing estimates of the indicator across assessment programmes yielded no systematic differences, and small differences were found for countries that participated in more than assessment programme. The proposed approach for reporting a single monitoring indicator is to report an indicator in five-year bounds by using an average of available estimates when MICS 6 data is not available. Estimates are proposed to be reported at the early primary, late primary; estimates at the secondary level are presented in this note for reference purposes.

## Background

**Mother-tongue education has been linked to improved learning outcomes.** Research suggests that mother-tongue instruction improves learning outcomes including the ability of children to learn languages later in school known as language transfer (e.g.: reviews by August & Shanahan 2006; Ball 2010; Piper et al. 2018). Immigrant children have been found to be less academically resilient and have lower cognitive achievement if they do not speak the language of instruction at home compared to immigrant children who do speak the language of instruction (OECD 2018). A number of studies have estimated country-level differences in cognitive achievement for ethnic minorities or indigenous peoples when their home language is not the language of instruction; these studies generally find lower performance on tests (e.g.: Stevens & Dworkin 2019).

**Few international studies on the prevalence of mother tongue instruction have been conducted; the latest round of MICS provides data to estimate SDG 4.5.2.** A review conducted by the UIS (2018) found one study that compared the number of languages of instruction to the number of languages spoken in countries (Kosonen 2017 in UIS 2018). The review found no instances of international household survey programmes that collected sufficient data for measuring SDG 4.5.2 except MICS 6 for which there was no data available at the time of their review. As of this draft, MICS 6 data is available for 20 countries, and UNICEF has provided UIS with estimates of SDG 4.5.2 for 13 countries (estimates were unable to be estimated for an additional 5 countries). MICS 6 asks children in school or who were in school about the language that their teachers used most of the time when teaching in addition to the language they speak most of the time at home. This permits an estimate of the percent of primary students learning in their home language. Of the international student assessments (TIMSS, PISA, LLECE, SACMEQ and PASEC) reviewed by the UIS (2018), only LLECE 2006 (SERCE) collected data on language of instruction that could be linked to students' responses on language used at home. All international assessment programmes collect data on whether the language of the test is spoken at home which is generally the language of instruction. UIS (2018) recommends using whether the language of test is spoken at home as a method to operationalize an indicator of SDG 4.5.2 using international student assessments.

**TIMSS and PISA collect data on whether the test language is spoken at home; PASEC 2014 and LLECE 2013 collect data on whether the dominant language of instruction is spoken at home.** For LLECE 2013, no question on language spoken at home is asked to students at the 3<sup>rd</sup> grade level; for 6<sup>th</sup> grade students, the question asks students to choose from a list of languages which they speak at home depending on the country including whether they speak Spanish or Portuguese (**Table 1**). For PASEC 2014, language spoken at home is similarly asked for both 2<sup>nd</sup> grade students and 6<sup>th</sup> grade students; for 2<sup>nd</sup> grade students, an enumerator asks the students. PASEC 2019 asks 6<sup>th</sup> grade students how often they speak the language of the test at home. SEA-PLM and PISA ask which language is spoken at home from a list of languages (which is coded as language of the test or other

in the dataset) as well as the frequency of using the test language at home with various family members. TIMSS and PIRLS ask how often the language of the test is used at home.

## Main methodological decisions

**1. Test language as a proxy for language of instruction:** As discussed above and by UIS (2018), language of instruction is generally not available in learning assessment data (with the exception of LLECE SERCE); however, the language of the test is generally intended to be the language of instruction. In absence of data on language of instruction, it is not possible to test whether the language of the test or the official language of instruction, even if the same, is actually what is used in the classroom. For example, in countries where the language of instruction is a *lingua franca*, it is possible that teachers may use local languages if children starting school are not sufficiently fluent in the *lingua franca*.

**2. Frequency of use of language at home:** PASEC 2014 and TIMSS 2015 ask students how often they use the language of the test at home, ranging from never to always. How should these be mapped to whether a child is deemed to be using the language of the test at home? LLECE 2013 and PISA 2018 ask simply which language of a list of languages (including the language of the test) is used most often at home. As a result, the frequencies reported by PASEC 2014 and TIMSS 2015, in order to be comparable with LLECE 2013 and PISA 2018 would need to reflect “most often”.

**3. Definition of measurement points:** The SDG 4.5.2 Indicator refers explicitly to primary school. At the primary school level, PASEC 2014 offers data on language at home for 2<sup>nd</sup> and 6<sup>th</sup> grade; LLECE 2013 and TIMSS 2015 offer data for 6<sup>th</sup> and 4<sup>th</sup> grade, respectively. Because research has found that the use of home language for learning to read in early grades is beneficial to children’s reading outcomes (e.g.: Melby-Lervåg & Lervåg 2011; Taylor & von Fintel 2016), distinguishing between early primary and late primary is proposed. For the assessments reviewed here, 2<sup>nd</sup> grade PASEC would be assigned to early primary; indicators derived from EGRA data would also be mapped to early primary to increase the sample size for early primary. TIMSS 2015 4<sup>th</sup> grade, LLECE and PASEC 6<sup>th</sup> grade would then be mapped to mid- to late-primary.

## Indicator methodology using learning assessment data

**Purpose:** Following UIS (2018), the proposed indicator for SDG 4.5.2 based on international learning assessments is the proportion of students that use the language of the test at home. This acts as a proxy for the proportion of students learning in their own language. SDG 4.5.2 explicitly states primary school students; however, this analysis included middle school students as well (TIMSS and PISA) for additional information.

**Definition:** 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). 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

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

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. **Table 1** lists the questions used in each assessment and how responses are mapped to 1 or 0.

**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	2 <sup>nd</sup> 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	6 <sup>th</sup> 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

**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)
PIRLS 2016/21 / TIMSS 2015 4 <sup>th</sup> Grade / TIMSS 2019 4 <sup>th</sup> Grade	Grade 4 Students (grades 2/3)	How often do you speak <language of test> at home? Fill one circle only.	<p>I always speak &lt;language of test&gt; at home</p> <p>I almost always speak &lt;language of test&gt; at home</p> <p>I sometimes speak &lt;language of test&gt; and sometimes speak another language at home</p> <p>I never speak &lt;language of test&gt; at home</p> <p>Mapped to yes if the child speaks the language of the test at home always or almost always; no if the language of the test is spoken at home sometimes or never.</p>
TIMSS 2015 8 <sup>th</sup> Grade / TIMSS 2019 8 <sup>th</sup> Grade	Grade 8 Students (end of lower secondary)	How often do you speak <language of test> at home?	<p>Always</p> <p>Almost always</p> <p>Sometimes</p> <p>Never</p> <p>Mapped to yes if the child speaks the language of the test at home always or almost always; no if the language of the test is spoken at home sometimes or never.</p>

**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 2015 & 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)	<p>"&lt;Language 1&gt;"  "&lt;Language 2&gt;"  "&lt;Language 3&gt;"  "&lt;...etc.&gt;"  "Other languages"</p> <p>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.</p>
SEA-PLM 2019	5 <sup>th</sup> 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.	<p>"&lt;Language 1&gt;"  "&lt;Language 2&gt;"  "&lt;Language 3&gt;"  "&lt;Language 4&gt;"  "&lt;Other language&gt;"</p> <p>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.</p>
PASEC 2019	6 <sup>th</sup> grade students	QE6_20. Parles-tu le français à la maison ?	<p>Toujours: yes  Souvent; Parfois; Jamais: no  no or invalid response: missing</p>



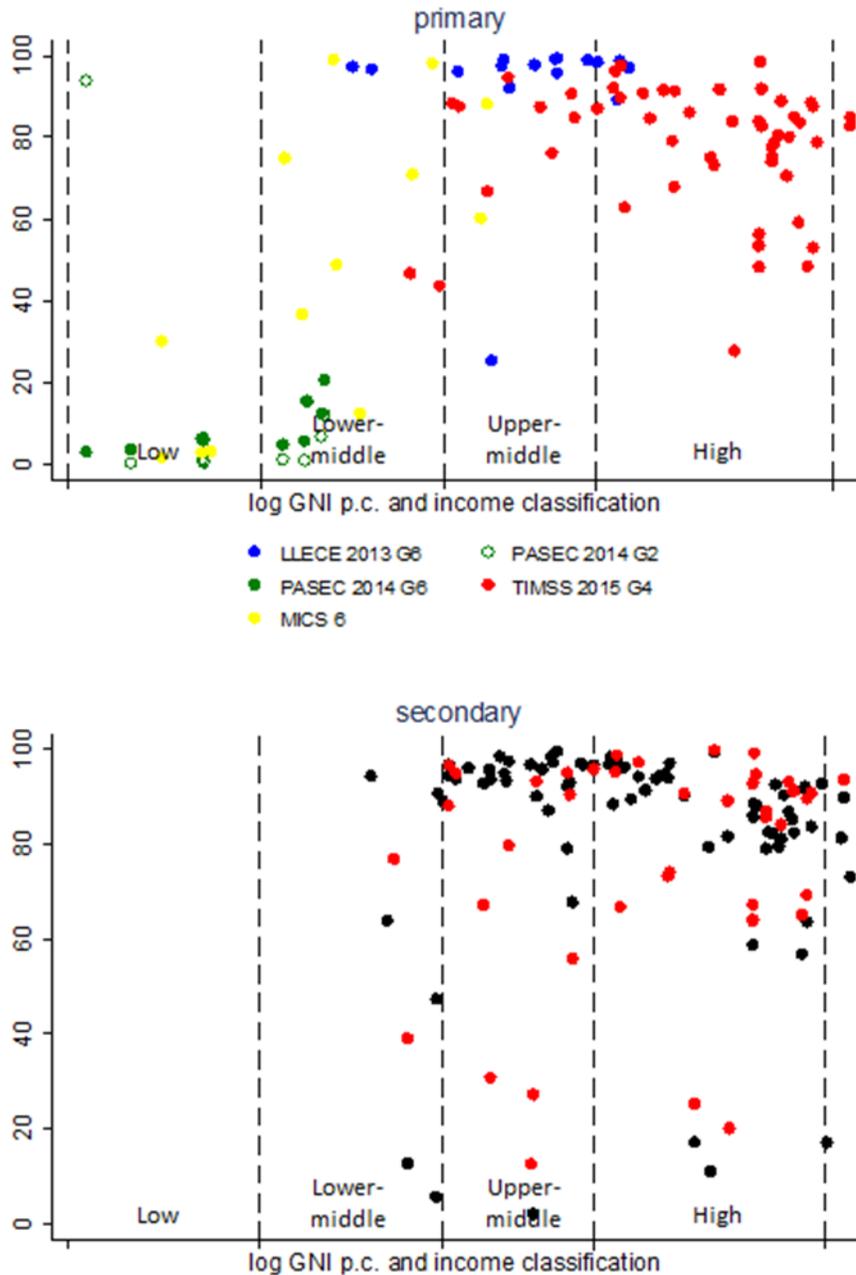
**Table 1. Data sources and questions on use of test language at home**

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)
-------------	-------------------	----------------------------	---

## Summary of estimated indicators globally

**The percent of students who speak the language of the test at home ranges from less than one percent to nearly 100 percent with most levels being above 87 percent.** At the primary level, the percent of 6<sup>th</sup> grade students in LLECE 2013 reporting using the language of the test at home was 89 percent or higher for all countries except Paraguay (**Figure 1**). Here, only 25 percent of students reported using the test language at home which may reflect the widespread use of Guaraní. PASEC reported the lowest percent of students using the test language at home. This is because the language of instruction, French (or English in some countries), is not commonly used at home. No more than 21 percent and 16 percent of students reported using the test language at home at the 6<sup>th</sup> grade level and 2<sup>nd</sup> grade level, respectively, with the exception of Burundi at the 2<sup>nd</sup> grade level which used Kirundi as the language of instruction and of the test. Burundi was also found to be outlier in terms of high performance in PASEC (PASEC 2015). For 4<sup>th</sup> grade TIMSS 2015, the percent of students using the test language at home ranges from 28 to 98 percent; the median for countries was 84 percent. For the 13 countries for which estimates using MICS 6 data is available, the percent of learning in their home language ranged from 1.9 percent to 99 percent; the median was 48.9 percent. At the secondary level, percentages of students using the test language at home were similar for PISA and 8<sup>th</sup> grade TIMSS. The lower range was 2 and 13 percent respectively, while the upper range was just under 100 percent for both. The median for PISA and 8<sup>th</sup> grade TIMSS was 91 and 89 percent, respectively.

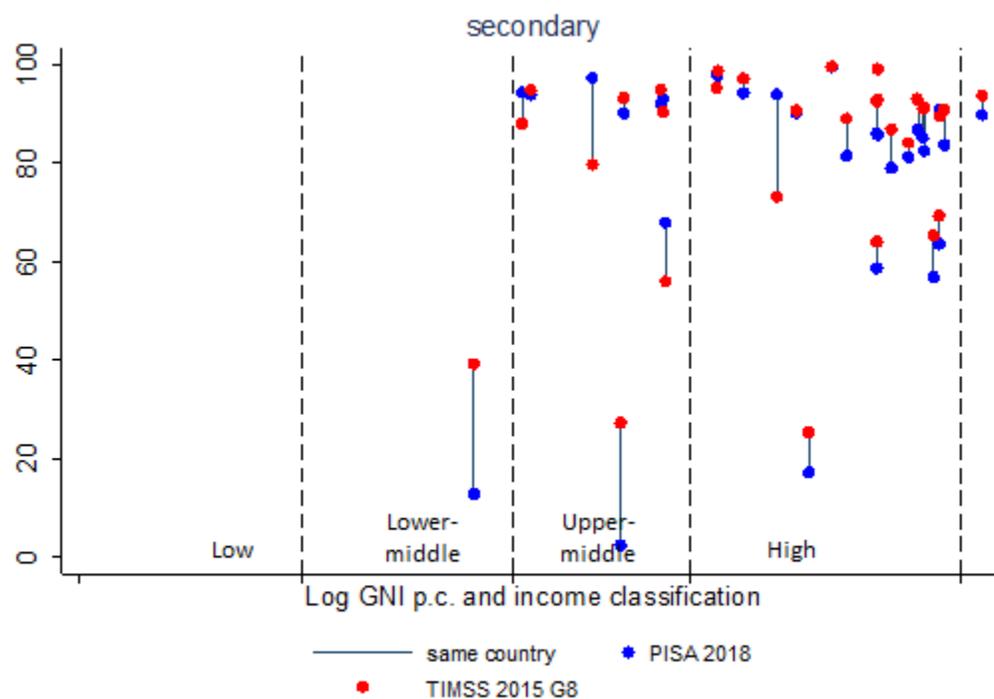
Figure 1. Percent of students whose language spoken at home is the same as the test language



Notes: MICS 6 presents estimates of the percent of primary students learning in their home language; for all other sources, estimates of the percent of students whose test language is their language at home. For LLECE and PASEC, the figures are the percent of students who speak Spanish or Portuguese at home and French or English (depending on the country), respectively. Source: author's calculations using LLECE, PASEC, TIMSS, and PISA data.

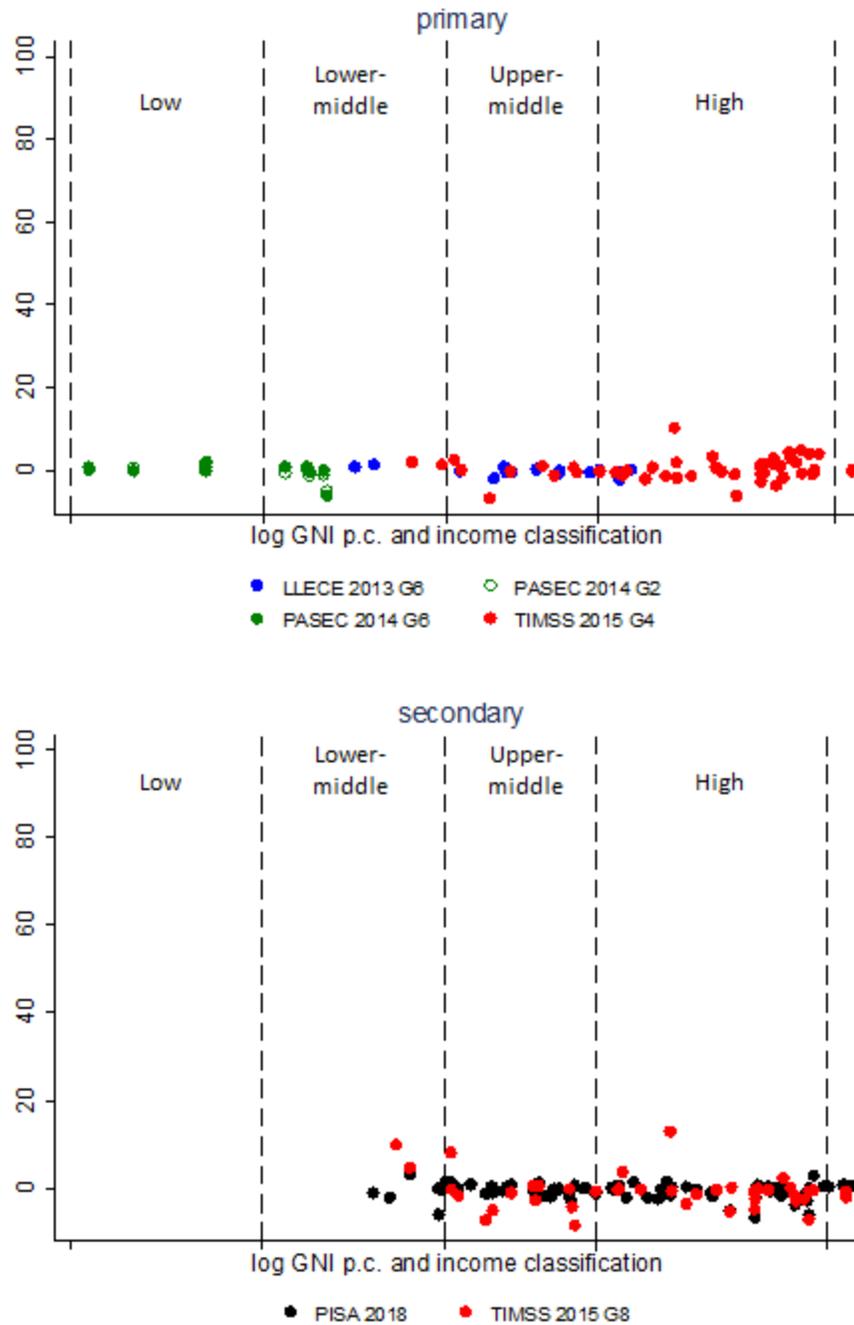
**No systematic differences between assessment programmes were found between countries that were sampled in both PISA 2018 and TIMSS 2015.** Only a few countries were sampled in more than one assessment or survey at the primary level. At the secondary level, 31 countries or locations had valid data for the percent of students that use the test language at home and were sampled in both PISA 2018 and TIMSS 2015 (only two countries were sampled in more than one assessment programme at the primary level). Comparing the differences in percentage of students using the test language at home between PISA 2018 and TIMSS 2015 reveals no systematic differences (**Figure 2**). In some cases estimates in TIMSS were higher than PISA and in some cases vice versa. Differences in percentages (in absolute value terms) ranged from 0 percentage points to 26 percent. The average difference was 6.8 percentage points.

**Figure 2. Percent of students who speak test language at home for countries sampled in both TIMSS 2015 and PISA 2018**



**On average, little difference in the percent of students using the test language at home was found between girls and boys.** Less than half of the 216 countries and locations included in the assessments had a gender difference in more than 0.8 percentage points (in absolute value terms), and 75 percent of countries had a gender difference of less than 2 percent. No systematic gender differences were found between assessment programmes (**Figure 3**). Estimates of SDG 4.5.2 from MICS 6 by gender are not currently available.

Figure 3. Percentage point gender difference in percent of students using the test language at home (male percent - female percent)



Source: author's calculations using LLECE, PISA and TIMSS data.

## Comparability across assessment programmes

**Reliability and comparability of assessment data estimates:** The questions asked in the four international assessments studied for this note asks similar questions allowing for comparable indicators across assessment programmes. PISA and LLECE ask which language the child speaks at home, most of the time, from a predefined list of answers. For PISA, this is reported in the data as whether the child speaks the test language at home most of the time or not. For LLECE, the first response is Spanish or Portuguese which are the test languages. TIMSS and PASEC ask child how frequently they speak the language of the test at home. As a result, by mapping these responses to “most of the time” (see Annexe Metadata), the indicators are largely comparable. This is reflected in few systematic differences between indicators measured for the same countries in TIMSS 2015 8<sup>th</sup> grade and PISA 2018.

## Proposed protocol for reporting the indicator

**Measurement points:** The indicator would be defined for three points of measurement given that PASEC offers data at two points during primary school: early primary would be defined as grades 1-3 while late primary would be defined after 4<sup>th</sup> grade. This division is designed to reflect research showing the importance of early grade reading in children’s home language as discussed above. Of the learning assessment data sources, 2<sup>nd</sup> grade PASEC would be mapped to early primary, 6<sup>th</sup> grade LLECE 2013, 6<sup>th</sup> grade PASEC 2014 and 4<sup>th</sup> grade TIMSS 2015, corresponding to mid to late primary, and PISA 2018 and 8<sup>th</sup> grade TIMSS 2015 corresponding to secondary school level (see **Table 2**). MICS 6 is also proposed to be reported at the early primary and mid- to late primary if sample sizes are sufficient (and confidence intervals reasonably small). If Early Grade Reading Assessment data were also added, these would be mapped based on their grade level (typically 1 to 3) accordingly.

**Table 2 Protocol for reporting percent of students learning in home language**

Measurement point (approx.)	Assessment	Grade	Target population
Early primary	PASEC 2014	2nd	Student in 2nd grade
Mid to late primary	LLECE 2013	6th	students in 6th grade
	PASEC 2014/19	6 <sup>th</sup>	students in 6th grade
	SEA-PLM 2019	5 <sup>th</sup>	Students in 5 <sup>th</sup> grade
	TIMSS 2015/19	4 <sup>th</sup>	students in 4th grade

Secondary	PISA 2015/18	n/a	15 year-olds in secondary
	TIMSS 2015/19	8th	students in 8th grade

Note that LLECE 2013 3rd Grade does not collect data on home language

**Monitoring indicators:** For the purposes of monitoring, a single estimate per country would be provided in five year bounds: 2011-2015, 2016-2020, 2021-2025, and 2026-2030. Because estimates of the indicator were largely similar between countries that were included in both 8<sup>th</sup> grade TIMSS and PISA, the proposed approach is to use an average of values available in each of these five year bounds. However, learning assessment data sources are proposed to be alternatives to MICS 6 (and subsequent rounds if estimates of SDG 4.5.2 remain possible) as these estimates most closely match the SDG definition.

**Reporting multiple estimates of the indicator-** Because of differences in the target populations between assessments within the proposed measurement points, publishing estimates from each country and assessment is proposed. This would aid researchers and provide transparency as to how the monitoring indicator was estimated.

## References

August, D. and T. Shanahan (2006). *Developing Literacy in Second-Language Learners: A Report of the National Literacy Panel on Language, Minority Children, and Youth*. Mahwah, N.J.: Lawrence Erlbaum Associates

Ball, J. (2010). *Enhancing learning of children from diverse language backgrounds: mother tongue-based bilingual or multilingual education in the early years*. Paris: UNESCO

PASEC (2015). *PASEC 2014 Education System Performance In Francophone Sub-Saharan Africa Competencies And Learning Factors In Primary Education*. Dakar: CONFEMEN

Melby-Lervåg, Monica, Arne Lervåg (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: a meta-analysis of the correlational evidence. *Journal of Research in Reading*, 34 (1), pp. 114-135

OECD (2018), *The Resilience of Students with an Immigrant Background: Factors that Shape Well-being*, OECD Reviews of Migrant Education, OECD Publishing, Paris, <https://doi.org/10.1787/9789264292093-en>.

Piper, Benjamin, Stephanie Simmons Zuilkowski, Dunston Kwayumba, Arbogast Oyanga (2018). Examining the secondary effects of mother-tongue literacy instruction in Kenya: Impacts on student learning in English, Kiswahili, and mathematics. *International Journal of Educational Development*. Vol. 59, March, pp: 110-127.

Stevens, Peter, Dworkin, A. Gary. (2019). *The Palgrave Handbook of Race and Ethnic Inequalities in Education*. Palgrave Macmillan

Taylor, Stephen, Marisa von Fintel (2016) Estimating the impact of language of instruction in South African primary schools: a fixed effects approach. *Economics of Education Review*, 50, pp. 75-8