



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



TECHNICAL
COOPERATION
GROUP



TCG4: Review of existing data sources to monitor SDG thematic indicator 4.5.2

TCG4/13

16-18 January 2018

Dusit Thani Dubai

133, Sheikh Zayed Road, Trade Centre,

Dubai, United Arab Emirates



International monitoring of language use in schools and instruction

A review of existing data sources to monitor SDG indicator 4.5.2

Table of Content

Table of Content	2
1 Introduction	5
1.1 Requirements for an international indicator	5
2 Overview of data sources	6
2.1 School surveys / Learning assessment studies.....	6
2.1.1 IEA-Studies: PIRLS, TIMSS, PrePIRLS.....	6
2.1.2 SACMEQ.....	7
2.1.3 PASEC.....	7
2.1.4 The Latin-American Laboratory for Assessment of the Quality of Education (LLECE)	8
2.1.5 EGRA	8
2.2 Household surveys	8
2.2.1 MICS	9
2.2.2 DHS	9
2.2.3 Living Standards Measurement Study (LSMS).....	9
2.3 National administrative data (EMIS)	10
2.4 Countries covered.....	10
3 Definitions and items used in major surveys	10
3.1 Language of instruction.....	11
3.1.1 Issues related to multilingual education policies (MLE)	11
3.1.2 Prevalence of multiple languages of instruction	12
3.1.3 Operationalization of language of instruction in different data sources.....	15
3.2 Operationalisations of language of student	18
3.2.1 How frequent speak students the language of instruction at home?.....	20
3.2.2 Are statements according to students' first language reliable?.....	22
3.2.3 Are statements according to student's home language reliable?.....	23
3.2.4 How consistent is reporting on first and home language when students report on both?	25
3.2.5 Details on specific languages	26
3.2.6 Operationalization of student's language	27
4 Sampling issues	32
4.1.1 Focus on grade or on level of education	32
4.1.2 Out of school children	34
4.1.3 Target populations and exclusions.....	35
4.1.4 IEA-Studies: PIRLS, TIMSS, PrePIRLS.....	35
4.1.5 SACMEQ.....	37
4.1.6 PASEC.....	37

4.1.7	The Latin-American Laboratory for Assessment of the Quality of Education (LLECE)	38
4.1.8	EGRA	38
4.1.9	MICS	39
5	Summary and Recommendations	41
5.1	Definition: Limit to main languages.....	41
5.2	Language of instruction.....	42
5.3	Definition: Home or first language (mother tongues, ...)	43
5.4	Frequency of language use	44
5.5	Inconsistencies in reporting by parents and children	45
5.6	Total primary education as target level.....	45
5.7	Data quality and precision	46
5.8	Detailed information on different languages	47
5.9	Comments on principles for Global Monitoring Indicators	47
	ANNEX A: Example questionnaire Items.....	49
5.10	PIRLS / TIMSS.....	49
5.11	SACMEQ.....	56
5.12	PASEC.....	57
5.13	LLECE: SERCE.....	58
5.14	EGRA	60
5.15	PISA (not administered at primary level, for reference purpose only).....	61
5.16	MICS.....	62
5.17	National Assessment Austria	63
5.18	Living Standards Measurement Study (LSMS)	64
	ANNEX B: Statistics and selected results.....	65
	ANNEX C: Participation status of countries for different surveys.....	78
6	References	84

Table of Figures

Figure 1: Share of students with other home language than language of test who are in schools that offer mother tongue reading instruction 14

Figure 2: Students report on the use of test language at home. 21

Figure 3: Comparison of two ways to ask for home language: Students who report to speak language of instruction always (or almost always) 22

Figure 4: Students and guardians report on the first language of the student: Dis-/agreement between respondents 23



Figure 5: Guardians report on the home language of students who say they speak language of test at home never or sometimes 25

Figure 6: Students report on the use of language of instruction at home and on having language of instruction as first language 26

Figure 7: Categories of pupils who spoke the language of instruction at home (SACMEQ III) 28

Figure 8: Percentage use of mother tongue as language of instruction in schools in Kenya and Uganda 34

Table of Tables

Table 1: Sample size in MICS 5, selected countries. 41

Table 2: Simulated sample size: For how many children would the child questionnaire be administered given the sampling procedure for MICS6 41

Annex B - Table 1: Share of students with other home language than language of test who are in schools that offer mother tongue reading instruction 65

Annex B - Table 2: Students report on the use of test language at home 68

Annex B - Table 3: Comparison of two ways to ask for home language: Students who report to speak language of instruction always (or almost always) (%) 70

Annex B - Table 4: Students and guardians report on the first language of the student: Dis-/agreement between respondents 72

Annex B - Table 5: Guardians report on the home language of students who say they speak language of test at home never or sometimes 74

Annex B - Table 6: Students report on the use of language of instruction at home and on having language of instruction as first language 76



1 Introduction

Following the adoption of the 2030 Agenda for Sustainable Development and the Education 2030 Framework for Action in 2015, the UNESCO Institute for Statistics has been leading the development of a thematic indicator framework for the follow-up and review of Sustainable Development Goal 4 on education. A set of 43 indicators, including the 11 global indicators recommended by the Inter-Agency and Expert Group on SDG indicators, was approved in October 2016 by the Technical Cooperation Group for SDG4-Education 2030 Indicators (TCG). At the same time, the TCG identified a number of indicators requiring further methodological development.

This paper aims to support the further development of the thematic indicator 4.5.2: *Percentage of students in primary education whose first or home language is the language of instruction* (UNESCO Institute for Statistics, 2017a).

The paper provides an overview of existing data sources containing information on languages of children and used in primary school instruction. Following an overview of the considered data sources, the paper discusses the following issues relevant for the monitoring of indicator 4.5.2:

- Which information is collected and which questionnaire items are used? Which definitions are in use for language of instruction and for students' languages?
- What are the target groups for reporting? Which implications do sampling issues have? What is the sampling precision (Sampling errors) of the data sources?
- Coverage of countries by Surveys?

The report includes a set of Annexes with more detailed material:

- ANNEX A: Examples for questions on languages from surveys.
- ANNEX B: Statistics and selected results.
- ANNEX C: Participation status of countries for different surveys.

A number of rough estimates concerning access to education in people's language exist already. For example, Kosonen (2017, p. 9) uses the distribution of languages spoken in the population of countries and calculates how many of them speak a language that is used at all for instruction. This paper does not aim to discuss all possible estimates but focuses only on data that cover children in primary education.

It is the nature of such a review, that it is critical about data definitions and data quality. The paper points at some general issues, that may not be able to easy overcome with current data. This should not suggest that current data sources could not be used for reporting or that it is here proposed to considerably change item sin the data sources. It must be considered that the sources are not under control of UIS and primarily designed for achievement assessment. Awareness of the issues raised here is needed, but a practical approach towards the indicator should sought, given the relevance of the indicator is judged high.

1.1 Requirements for an international indicator

The recommendations in this paper are oriented on a major idea guiding UIS in the development of indicators: Every attempt should be made to use existing data sources rather than conduct new data



collections in order to manage respondent burden and ensure efficiency of the UIS statistical system (UNESCO Institute for Statistics, 2017b). Therefore, the recommendations do not aim to propose new approaches to monitor 4.5.2. They focus for existing material on aspects to consider when aiming to translate different data sources into one framework of reporting. From that guidance can be derived for future surveys (including national surveys) to better match existing data.

When considering the creation of new SDG indicators, the principles for setting SDG indicators should be considered (Leadership Council of the Sustainable Development Solutions Network, 2015, p. 17), although the principles may not apply as strict for thematic indicators. The principles are:

1. Limited in number and globally harmonized
2. Simple, single - variable indicators, with straightforward policy implications
3. Allow for high frequency monitoring
4. Consensus based, in line with international standards and system - based information
5. Constructed from well - established data sources
6. Disaggregated
7. Universal
8. Mainly outcome - focused
9. Science - based and forward - looking
10. A proxy for broader issues or conditions

The summarizing discussion at the end of this paper will comment for the indicator 4.5.2 on the 10 principles.

2 Overview of data sources

This section provides an overview of the data sources considered in this report. There are two groups of data sources to be distinguished: i) Primary school surveys, typically assessment surveys, which collect data on pupils within their schools, for selected grades; ii) Household surveys, which collect data from a sample of households with children from all ages and primary grades.

2.1 School surveys / Learning assessment studies

The school surveys and learning assessment studies discussed here are sample survey of primary schools designed to measure student achievement. The sampling units are schools and typically classes within schools. The design of the studies is optimized to monitor learning outcomes at critical steps in the school careers. They are typically conducted towards the end of primary education to summarize learning outcomes although some surveys collect also data for earlier grades.

2.1.1 IEA-Studies: PIRLS, TIMSS, PrePIRLS

The International Association for the Evaluation of Educational Achievement (IEA) conducts international large scale student assessments since more than 5 decades. Reporting on students having a different home language than the language of instruction is a constant element of reporting in IEA studies for many years (Elley, 1992, p. 59). The Trends in International Mathematics and Science



Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) are conducted at the primary level.

At the primary level PIRLS collects data on reading comprehension every 5 years, TIMSS on mathematics and science every 4 years. PIRLS is designed to complement IEA's TIMSS assessments of mathematics and science at the fourth grade (Mullis & Martin, 2015). Thus, both studies, coordinated by the TIMSS/PIRLS International Study Center (Boston College), follow very similar samples, procedures and methodologies and are mainly based on consistent questionnaire frameworks (Hooper, Mullis, & Martin, 2015; Hooper, Mullis, & Martin, 2013). This generally allows to pool data from items on language of students from both surveys to maximize coverage. PrePIRLS complements PIRLS, but is designed to test more basic skills that are prerequisites for success in reading comprehension. PrePIRLS collects the same background data as PIRLS and can thus supplement PIRLS. Countries participating in PrePIRLS test sometimes other grade than the 4th grade.

This paper discusses the IEA survey projects as one. TIMSS/PIRLS assesses pupils in grade 4 and thus provides information with respect to higher grades of primary education. As for all data from student assessment projects, it must be kept in mind that the purpose of background data in PIRLS and TIMSS is to study the factors associated with children's reading, mathematics or science achievement (Martin, Mullis, & Foy, 2015, p. 67).

From IEA studies data on grade four are available for eight reference years: 1995, 2001, 2003, 2006, 2007, 2011, 2015 and 2016. The next studies are scheduled for 2019 and 2021. Data are available from at least one study for 65 countries. In the recent round (TIMSS 2015/PIRLS 2016) 47 countries have data on language of students for grade 4. Some countries or subnational entities participate in PIRLS/TIMSS or PrePIRLS with other grades than grade 4: Malta, South Africa¹, New Zealand and England with grade 5 and Botswana, Honduras and Kuwait with grade 6. (2011).

2.1.2 SACMEQ

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) is a collaborative network of 16 ministries of education in Anglophone East Africa. SACMEQ projects are large scale assessments in literacy, and numeracy. The first SACMEQ assessments were (1995 to 1999) were in the tradition of the IEA assessments. SACMEQ is administered in school to children in the 6th grade of formal school. Up to know four rounds have been carried out: 1995-1999, 2000-2004, 2006-2011 and 2011-2014. Data and material of the last round (SACMEQ 4) have not yet been published, this review draws on material from SACMEQ 2 and 3. SACMEQ administers background questionnaires for students, teachers and school heads. Data for 16 countries are available from SACMEQ.

2.1.3 PASEC

PASEC (Analysis Programme of the CONFEMEN Education Systems) is a assessment project in 13 countries in Francophone West Africa and three in South East Asia. PASEC assesses student achievement in mathematics and reading French since 1993. The program is managed by CONFEMEN,

¹ South Africa tested grade 5 with PIRLS and grade 4 with PrePIRLS.



the conference of education ministries in francophone countries. PASEC is typically administered to students in 2nd and 5th grades. PASEC was conducted in three periods of work (1991-201, 2011,2012, 2014). PASEC administers background questionnaires for students, teachers and school heads. 10 Countries participated in PASEC 2014 and 15 countries joined PASEC 2019.

2.1.4 The Latin-American Laboratory for Assessment of the Quality of Education (LLECE)

The LLECE (Latin-American Laboratory for Assessment of the Quality of Education) is the Latin American network for the assessment of education quality. It is coordinated by UNESCO's Regional Bureau for Education in Latin America and the Caribbean (OREALC). Three assessments were conducted: PERCE (1997) started with the assessment of reading and mathematics, SERCE (2006) extended the areas to assessed Reading, Writing, Mathematics and Sciences and was conducted in 16 countries. The most recent round of LLECE, TERCE (2013) was conducted in 15 countries in the region. LLECE assesses students in the third and sixth grade.

2.1.5 EGRA

The Early Grade Reading Assessment (EGRA) and the Early Grade Mathematics Assessment (EGMA) models are designed to provide simple, low-cost measures of literacy and numeracy for early grades (EPDC). Development and implementation of EGRA begun in 2006 by RTI International under the USAID EdData II project. Both assessments are oral. EGRA has been administered in at least 11 countries and 19 languages under the EdData II program and has been used in more than 30 countries and 60 languages by other organizations. The EGRA instrument is designed to be a sample-based "system diagnostic" measure. Its purpose is to document student performance on early grade reading skills in order to inform ministries and donors regarding system needs for improving instruction (RTI International, 2009). EGRA documentation emphasises that it should not be used to compare results across languages. Yet, this seems not applicable for background questionnaire items, if background items are used identical in several assessments. EGRA presents itself as toolkit and its implementation seems country specific, in some cases in order to undertake evaluations of interventions. It seems that the EGRA sample is less strict in the implementation of national representative samples than other survey programmes.

2.2 Household surveys

International household surveys are designed to monitor many aspects of the lives of children and families mainly in developing countries. The surveys focus on those issues that directly affect the lives of children and women, including health and wellbeing. The surveys are not explicitly designed to monitor education, but information on education, especially attainment and participation, are standard elements of the surveys. Yet, in the past no information was collected that could be used for indicator 4.5.2. This is only envisaged for the coming round of MICS.



2.2.1 MICS

UNICEF's Multiple Indicator Cluster Surveys (MICS) was conducted since 1995 and became the largest source of statistically sound and internationally comparable data on women and children worldwide. In the last two decades about 300 MICS surveys were conducted in 108 countries. In the first five rounds of MICS education was covered by questions concerning participation, but no information was gathered on the instruction of children.

Relevant for the monitoring of indicator 4.5.2 will be the sixth round of MICS, that is currently in the field. The MICS instruments are now complemented by a special questionnaire for children aged 5 to 17 and thus focusing more on school aged children (UNICEF-MICS, 2017). In this questionnaire the language of children and their instruction is added as specific item. It is announced, that 41 countries are scheduled to participate in MICS6². The questionnaire, administered to children aged 5 to 17, will provide a sample of children from all grades. One child in the target age group, randomly selected, is interviewed in each household. Since all children, in and out of school, are included in the sample, analysis can be complemented by looking at the language of out of school children.

The new module for children is one of several modules. Not all countries administer all modules. The MICS recommendation is to keep applicable modules and questions that will collect information identified during data needs assessment process as non-existing, in need for more disaggregation or that are based on internationally comparable methodologies. In the beginning MICS6 round, particularly because of the need for data toward SDGs monitoring, UNICEF did not observe lot of reduction in specific questionnaires. Preliminary information indicates wide use of the children module, although not all countries and world regions have yet entered the survey design state and information is preliminary. It can be expected that many but not all of the 41 countries the joined in for MICS 6 will cover the questionnaire for school age children and thus collect data relevant for 4.5.2.

2.2.2 DHS

The DHS Survey currently does not include items relevant for the monitoring of languages of instruction (DHS). It is not further discussed in this review.

2.2.3 Living Standards Measurement Study (LSMS)

The Living Standards Measurement Study (LSMS) surveys have been conducted in dozens of countries around the world. The survey design is unusual with respect to the questionnaires used. Besides questionnaires for households also other levels of respondents are considered, most common is the community questionnaire. A small number of countries also use school questionnaires that can be linked to the household responses.³ The household questionnaires allow for identification of the languages used. Grosh and Glewwe (2000) provide an example questionnaire for schools including a question on language of instruction. An actual application of the example language question could, however, not be identified.

² <http://mics.unicef.org/surveys>

³ To find examples see <http://iresearch.worldbank.org/lsmsslmsurveyfinder.htm>



2.3 National administrative data (EMIS)

A small but growing number of countries cover student language in their national education statistics. Education census and other routine administrative data collections (commonly referred as “EMIS”) could also serve as national data sources to monitor instruction in the home/first language. Most EMIS data collections are done at the school or class level in the form of school aggregated data. In such systems the collection of student level characteristics is currently not common, but it exists. The EMIS system of Scotland collects aggregated data from schools on the language of students. But it appears, that this is possible because the system relies on school management systems that host student level data (Scottish Government [ScotXed], 2017). Also, some German states collect language of students as aggregated data (Kemper, 2017).

Other countries build up student registries which include student level information. Mellander (2016) sees register data as a fundamentally Nordic (North European) phenomenon. Yet, other countries have started building up similar systems. Mellander reports that data on the mother tongue of pupils are a common feature of Nordic registers. Unfortunately, this could not be confirmed by documents of the responsible statistical offices.

A special form of collection of national student level data are student/parent surveys outside the context of assessment studies. In some countries they are conducted by schools in order to gather data that are used to complete the school census. Some systems administer such surveys centrally. The Australian state of Victoria used a student survey with the school census. Parents are obliged to report on their family’s characteristics (e.g. parental education, occupation, country of birth) including home language. (Department of Education and Early Childhood Development, 2013). In a similar way, the education system in Flanders (Belgium) collects data from parents (Flemish Department of Education and Training, 2015, p. 29). In England, parents are obliged to participate in the school census and report on language (Department for Education, 2017). The German state of Hamburg collects data on the social background of students by a survey of all schools but only a sample of students within the schools (Schulte, Hartig, & Pietsch, 2016) Student characteristics are thus available for all schools and even though not as census, the data are highly representative for the system.

2.4 Countries covered

The reviewed international projects cover a wide range of countries in primary education worldwide. In all world regions. 50 countries are listed alone as participants of PIRLS/TIMSS when combining data from the 2015, 2016 and forthcoming 2017 assessment. Taking all school assessment surveys since 2010 together the number of participating countries almost doubles to 97 countries. Assuming that almost all of the countries in MICS6 administer the questionnaire for children aged 5 to 17, the number of countries could reach up to 120 countries. For details see Annex.

3 Definitions and items used in major surveys

This section describes how the different data sources capture the information needed for 4.5.2: language of instruction and home/first language of the students. It starts for the two pieces of information with a discussion how the definition of 4.5.2 can be operationalized.

3.1 Language of instruction

The proposed label for the indicator 4.5.2 is *“Percentage of students in primary education whose first or home language is the language of instruction.”* This wording suggests that one language can be identified as main language of instruction. For the development of a global indicator it is crucial to clearly define, whether the purpose of the indicator is to report on students that have (some) instruction in their home/first language or to limit it to report on students that have their home/first language as the main language that is used most of the time.

Most data sources reviewed assume a single formal language of instruction. School assessment surveys seem to assume that this is a fixed property of the class rooms or tested groups. Only EGRA includes an approach to capture languages of instruction beyond formal or official settings (see Questionnaire example 23 in Annex). Gove and Wetterberg (2011, p. 155) used a modified version of “Stallings snapshot tool” to capture real class room observations on the language choices used in Kenya and Uganda. Surely this complex and demanding approach cannot serve as a model for international monitoring, but it demonstrates that a presumably dichotomous variable can be much more variant when closely observed.

However, the perspective that language of instruction can be captured as one single and fixed language would be considered as falling short of class room realities by many researchers. Kosonen (2017, p. 16) promotes the use of more complex questionnaires to determine the language of instruction at school level. UNESCO’s Multilingual Education Mapping Exercise Questionnaire⁴ illustrates the complexity of approaches capturing language of instruction beyond the assumption of a single fixed language. The need for such approaches for research and local level education policies should not be questioned, for indicators at a regional or global level appear such approaches too ambitious.

It should also be discussed, whether an indicator that is strongly geared towards a monolingual perspective on language of instruction (and students home language as well) is in line with policy statements on SDG goals. Key messages that a certain percentage of children do not have access to education in a language they speak can not be backed up by indicators that are blind towards bilingual education beyond the main language of instruction. Also stating that certain languages are not spoken in class room cannot be derived from a monolingual perspective on the main language of instruction. In my view it is questionable whether GEM messages and recommendations, as given in the recent GEM (UNESCO, 2016, 1,9) would be in line with an indicator focusing on single main languages.

3.1.1 Issues related to multilingual education policies (MLE)

If the definition for 4.5.2 is clarified towards a focus on ONE main language of instruction, this has major consequences in the interpretation of the data. Many countries tackle the issue of language of instruction by implementing multilingual education policies (MLE). UNESCO (2003) argues also for multilingual education: “To reduce the disadvantages experienced in school by children of a language background that differs from the language of instruction, mother-tongue based bilingual and

4

<http://www.unescobkk.org/education/multilingual%E2%80%93education/mle%E2%80%93mapping%E2%80%93data>



multilingual education is broadly considered to be the best solution. In bilingual education, both the mother tongue and a regional or national language is used in education. In multilingual education, besides these two languages also at least one international language is used". It is unclear how simple indicators can capture gradual progress towards this desired situation.

Typical for MLE is that instruction is provided in the first/home language and in another language (national language, common language, trade language, colonial language. etc), which is also the language of instruction at further education. In early grades the home/first language is the dominant language of instruction, but in higher grades instruction is shifted to be mainly in the other language. This is the case in South-East-Asia (e.g. Cambodia, Philippines, Timor-Leste) (Kosonen, 2017, 5,6,8), Sub-Saharan Africa (e.g. Ghana (Ministry of Education Ghana, 2016, p. 1), Guinea, Gambia, Madagascar, South Africa, Swaziland (Albaugh, 2012). The shift towards other languages in instruction is commonly after three or four grades. The shift between languages as main language of instruction might be at different times for different subjects, as in Mauritania: Arabic is the language of instruction for all subjects in grade one. French is introduced for a few hours a week in grade two. In grade three, French is used to teach math, and in grade five, French is introduced to teach science. Arabic is used for all other subjects (Albaugh, 2012, p. 42). In this case, a definition focusing on one main language of instruction in a tested subject may lead to different interpretation dependent whether instruction in math or science or reading would be in focus of the survey.

When monitoring progress in implementation of MLE, there is a risk that 4.5.2 is not sensitive to changes as long as a home languages does not replace the previous dominant language as main mean of instruction.

The use of home languages as second but not main language of instruction is also relevant for many systems with large migrant populations. Although there is no clear consensus in the research whether for migrant students bilingual education is more effective than using the language of mainstream instructions (Herzog-Punzenberger & Schnell, 2012, p. 239; Nusche, 2009, p. 29), most developed countries provide some mother tongue education for big numbers of their migrant students. For more details see below.

3.1.2 Prevalence of multiple languages of instruction

The data from PIRLS 2011 allow to judge how many students have potentially access to reading instruction in their home language beyond the instruction in the tested language. School heads are asked: "*For students in <fourth grade> and below, does your school make provisions for reading instruction in mother tongue for students whose mother tongue is not <language of test>?*"⁵ For the analysis also countries participating in TIMSS 2011 but not in PIRLS 2011 and countries participating in PrePIRLS 2011 were also included, since the studies shared a common questionnaire.

Figure 1 shows that a substantial share of students having a different home language have potentially access to reading instruction in their languages at their school despite having not their language as

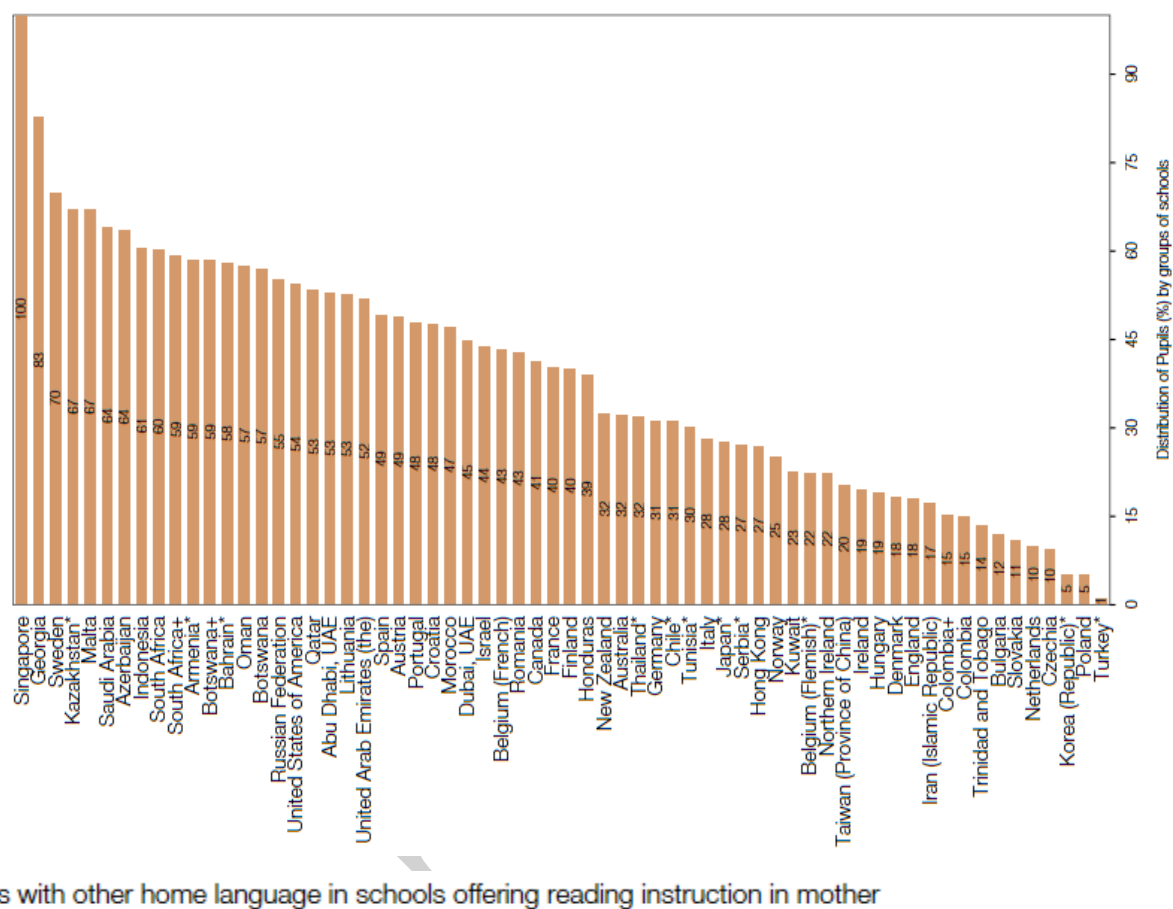
⁵ The phrases <fourth grade> and <language of test> are placeholder for country specific text. Countries replace it with the appropriate national name for the grade tested and with the name of the test-language. Some countries have as many versions of the questionnaire as they use different languages of test. See Questionnaire example 13 in the Annex.



main instruction language. In one third of countries more than half of the grade 4 students with other home language are enrolled in schools that offer mother tongue reading instruction for students whose mother tongue is not <language of test>. In Singapore, Georgia, Sweden Kazakhstan, Malta this is the case for 2/3 of those students. Only 14 countries have less than one fifth of grade 4 students with other home language enrolled in schools offering mother tongue reading instruction: Ireland, Hungary, Denmark, England, Iran, Colombia, Trinidad and Tobago, Bulgaria, Slovakia, Netherlands, Czechia, Korea (Republic), Poland, Turkey.

Even though PIRLS data do not allow to judge the intensity of instruction in those second languages, the figure shows that bilingual instruction settings are wide spread. If students in grade 4 would simply be classified by their current main language of instruction (in the case of PIRLS the language of test) the efforts towards bilingual reading instruction would not be reflected in indicator 4.5.2. The case of Singapore illustrates problems in interpretation of data. Here 68% of the students tested in TIMSS 2001 indicate a different home language as the language of test (language of instruction)⁶. Yet, all of them are in schools that offer reading instruction in the mother tongue. This remarkable picture is related to MLE in Singapore. In school, English is the medium of instruction for all academic subjects at all grade levels (including mathematics and science), except for the mother tongue language subjects. A fundamental feature of Singapore's post-independence education system is its bilingual policy. Students are encouraged to be proficient in both English and their own mother tongue language (Malay, Chinese, or Tamil). In 2010, 71 percent of the literate resident population in Singapore was literate in two or more languages, an increase from 56 percent in 2000. (Mullis, Martin, Goh, & Cotter, Singapore). A similar complex picture discusses Gove and Wetterberg (2011, p. 163,164) for Kenya and Uganda. Although about 45% of instruction time in combined grade 1 to 3 are in mother tongue in the tested Kenyan regions, an analysis separated by subject indicates a use of mother tongue language in less than 14% in most subjects, except, not surprisingly, in mother tongue instruction.

⁶ Because the data are collected with TIMSS, the language of instruction is set to be the language of science and mathematics. Having the same assessment with PIRLS would show different results, because reported relative to the language of instruction of reading instruction.



* Country participated in TIMSS 2011 but not in PIRLS 2011.

+ Country participated in PrePIRLS 2011.

Some countries are listed twice with different grades for PrePIRLS and PIRLS."

The school questionnaires of TIMSS 2011 and PrePIRLS 2011 are identical to PIRLS 2001.

Source: PIRLS2011, TIMSS 2011, PrePIRLS2011. Authors own calculations. See Annex X.

Figure 1: Share of students with other home language than language of test who are in schools that offer mother tongue reading instruction

Reporting on access of students to instruction in their home language with a limited focus on one main language of instruction would in countries with big migrant populations most certainly lead to conflicting messages. If monitoring recognizes access to education in the students language only if their language is the main language of instruction, indicators could only show increasing percentages if migrant students are segregated in classes with their own language used as main language for instruction. Yet, ethnic, social or language segregation is seen as one of the issues education systems should reduce, there is little evidence in the effectiveness of pull out classes with instruction in the language of migrants (Nusche, 2009, p. 9,27). In education systems with large migrant population, the diversity of languages is increasing. In order to provide mother tongue education to migrant students



a large number of languages would need to be addressed. Even the biggest language groups amongst migrant students form typically a minority amongst the students with other home languages. E.g. in Austria speak 24% of all primary students a language different from the main language of instruction (German). Yet the biggest groups amongst them, students speaking Turkish, Bosnian, Croatian or Serbian account only for half of the non-German speaking students. In urban Vienna 28% of all students speak languages different from German, Turkish, Bosnian, Croatian or Serbian (Bruneforth, Lassnigg, Vogtenhuber, Schreiner, & Breit, 2016). Language policies that succeed to offer these languages as main language of instruction to them see neither realistic nor desirable.

3.1.3 Operationalization of language of instruction in different data sources

3.1.3.1 IEA-Studies: PIRLS, TIMSS, PreTIMSS

For IEA studies the language(s) of instruction for tested students are not explicitly collected from schools, teachers or students. Instead the concept of “language of test” is used. Information on which language is to be used for the test is collected during the sampling process. Language of test is used as synonym for language of instruction at the given grade, since it is given that students tested in the given “language of test” received substantial instruction in that language in the tested subject. Therefore, it seems for the given sample a valid operationalization of language of instruction. IEA reports on the official languages of instruction and shows the match with the test languages. In almost all most countries there is a match, few countries have more official instructional languages as used for testing (Mullis, Martin, Kennedy, & Foy, 2007, p. 288).

Issues arise rather in the context of sampling, as discussed below, since the tested sample is in some countries restricted to the languages of test used and sample exclusions are in many countries related to issues of languages.

In the PIRLS/TIMSS/PrePIRLS-Round 2011 40 out of 61 countries used one single language of test for their assessment and thus reflected one single language of instruction in the data. Less than one third of countries (17) use two languages for testing: Bahrain, Canada, Finland, Hong Kong, Ireland, Israel, Italy, Kazakhstan, New Zealand, Northern Ireland, Norway, Oman, Qatar, Romania, Saudi Arabia, Slovakia, the United Arab Emirates. Spain used 4 languages in TIMSS 2011 and 5 in PIRLS 2011⁷. South Africa represents 11 languages of test in the data of prePIRLS for grade 4⁸, but just 2 languages when testing PIRLS at grade 5.

The limitation that only tested languages show up as language of instruction in TIMSS or PIRLS does not preclude, that in several countries also small language groups are covered. The smallest groups tested with their language of instructions are, unweighted, between 19 and 150 students in PIRLS: Spain, Italy, Saudi Arabia and New Zealand. Four more countries cover languages with also less than 1000 students in the sample.

⁷ Castilian, Basque, Catalan, Galician, Valencian.

⁸ Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Sepedi (Sesotho sa Leboa), Sesotho, Setswana, siSwati, Tshivenda, Xitsonga. Each language was tested with more than 1000 students in the sample of about 15,000 students in of PrePIRLS2011.



PIRLS/TIMSS Background questionnaires give in addition some information on the use of languages of instruction at school. In PIRLS 2011 (IEA, 2013) school heads are asked “*For students in <fourth-grade> and below, does your school make provisions for reading instruction in mother tongue for students whose mother tongue is not <language of test>?*”. The data provide some insight in the availability of instruction in other languages. For data see Figure 1. Unfortunately, PIRLS 2016 discontinued this item (IEA, 2016).

3.1.3.2 SACMEQ

The SACMEQ project was initially modelled after IEAs Reading Literacy Study 1991. To the authors knowledge, processes are therefore similar to those used in IEA studies. From a review of materials on SACMEQ 2 and SACMEQ 3 it appears that SACMEQ also operationalizes the language of instruction as “language of test”, which is mainly English for grade 6 in the participating countries. In SACMEQ 3 (SACMEQ, 2007) there are question home language of students, but no questions on language of instruction. The questionnaire published uses always “English” as term for language of instruction. It can be assumed that there is, as for IEA studies, a main language of instruction considered.

As discussed above, due to the focus on grade 6, the data arising from SACMEQ would hide the use of local languages in early grades of primary education. This is an artefact of testing and the grade chosen. SACMEQ (p. 20) reports on the decision about the grade to test for SACMEQ 3 “the NRCs considered that testing pupils at grade levels lower than Grade 6 was problematic – because in some SACMEQ countries the lower grades were too close to the transition point between the use of local and national languages by teachers in the classroom. This transition point generally occurred at around Grade 3 level – but in some rural areas of some countries it was thought to be as high as Grade 4 level.” Since the questionnaires for SACMEQ 3 do not include items on language of instruction in lower grades during the students’ school carrier, SACMEQ data can only report on language use towards the end of primary education.

3.1.3.3 PASEC

The PASEC assessment is administered in the official language of instruction. This is typically French, also exceptions exist as is the case of Kirundi in Burundi at the beginning of the primary cycle, and English in Anglophone Cameroon (Programme d’Analyse des Systèmes Educatifs de la CONFEMEN, 2015, p. 20). PASEC collects information on the language of instruction directly from the teacher. It asks them how frequent they use the mother tongue of children in their instructions: Almost always, often, sometimes or nearly never. The research for this paper found no analysis using this data.

3.1.3.4 The Latin-American Laboratory for Assessment of the Quality of Education (LLECE)

The LLECE studies are conducted in Spanish and Portuguese language. As for the above discussed surveys, the language of instruction relevant for the analysis of the data is language of test. Yet, different from the above discussed surveys, LLECE collects directly information on the languages spoken in instruction for grade 3 and 6. Teachers for grade 3 and 6 are asked which language they use most frequent in class. (see Questionnaire example 18 in Annex). In addition. students at grade 6 are asked which language the teacher uses most in class (see Questionnaire example 19). The language categories are identical for language used in class (language of instruction) and home/first



language of students. This allows a match of languages for the main language of instruction, even though bilingual instruction, so existing, would not be captured. The language categories differentiate types of languages as well as languages: Spanish/Portuguese, Foreign language and indigenous languages.

3.1.3.5 *Early Grade Reading Assessment (EGRA)*

The implementation of EGRA differs between countries where it is implemented. In some countries EGRA is only conducted for specific language of instruction and schools are selected based on the languages used (e.g. monolingual in Myanmar (WorldBank, 2015) or in French and Wolof in Senegal (Sprengel, 2008)). In other countries several languages of test are used without prior knowledge of the language used in a given school. EGRA collects, at least in some surveys the information on language of instruction from head teachers. "To ascertain what language was typically used as the language of instruction by teachers ... the head teacher at each school was asked to indicate the language (Haitian Creole or French) that was used by Grade 1 and 2 teachers in that school to teach mathematics" (RTI International, 2014b, p. 17).

Yet, the use of different languages reported by teachers is not necessarily an indication of a policy of language of instruction but can simply indicate a limitation in the command of languages by the teacher him/herself, as RTI International (2014a, p. 40) discusses for the case of Morocco.

3.1.3.6 *MICS*

In the case of MICS, as with any household survey, information on education is collected outside educational settings. All information is gathered from children and their guardians directly. In that sense, language of instruction, is an independent variable from the survey processes. For the forthcoming MICS6 data on language of instruction of children aged 5 to 17 will be collected for the first time. The children will be interviewed (in the presence of their mother's/caretaker's) and will be asked what language their teachers use most of the time when teaching them in class. The list of language options is identical to the list given when asking for the home language. The interviewer should select the most appropriate languages from a list and include these in the question to the child (see Questionnaire example 25 in Annex).

3.1.3.7 *Living Standards Measurement Study (LSMS)*

(Grosh & Glewwe, 2000, Vol. 3, p. 37/op1) proposes for the LSMS the use of school questionnaires. An example questionnaire for schools including a question on language of instruction exists. The item used to identify the language of instruction is different from other examples in that multiple language settings are reflected. The person interviewed for the school is asked for the language used in mathematics instruction, reading instruction and general conversation. See Questionnaire example 28 in Annex. An actual application of the example language could not be identified at <http://iresearch.worldbank.org/lsmssurveyfinder.htm>.



3.1.3.8 National administrative data (EMIS)

All examples of national administrative data sources reviewed focus only on the language of the student. From the available examples of how the data are used in national reporting, the focus lies on the students and not on schools and instruction. All national statistical reports the author could find report on the distribution of languages spoken by pupils but not in relation to the language of instruction.

It appears that language of instruction is either perceived as static, meaning identical to the main language of a country, or as a characteristic of a given school or educational programme. In the Austrian student register all students are assigned to educational programmes (*Schulform*). The national list of educational programmes allows the identification of programmes with provision in languages other than German, the main language in Austria (Bundesministerium für Bildung, 2016). This allows the determination of the language of instruction for students in schools catering for linguistic minorities. Yet, the system does not allow for a clear distinction between schools with instruction in the minority language only or with bi-lingual instruction. Language of instruction itself is not a variable.

3.1.3.9 Census and national household surveys

School and instruction are not topics traditionally covered in population censuses. No examples including questions on instruction could be found. For household surveys (UNESCO Bangkok, 2008, p. 17) proposes to capture the language of instruction within the context of the literacy module of certain household surveys. Here the question "In what language did (name) first learn to read and write?" is intended to capture the language of instruction in the school the respondent visited.⁹ The problem with the module is, that it targets only persons aged 15 and above. Thus resulting data could only inform retrospectively on language of instruction. Country examples for the application of the proposed module could not be found.

3.2 Operationalisations of language of student

The proposed label for the indicator 4.5.2 is "*Percentage of students in primary education whose first or home language is the language of instruction.*"

UNESCO Institute for Statistics (2017a, p. 50) defines further: *First or home language is defined 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.*

This wording suggests that the label "first language" or "home language" are exchangeable or that at least an aggregation of data sources is open to allow a mix of the two.

Secondly the wording suggests that children are monolingual, meaning they have one main language, which can be set in relation to the language of instruction. This is clearly confirmed by UNESCO Institute for Statistics (2017a, p. 50), although the purpose of the indicator is described as "To measure

⁹ http://old.unesco.kz/education/2007/CARK_EFA_MDA/day4/statistic/Litproj_en.pdf



the extent to which children in primary education are learning in a language with which they are familiar and in which they are likely to be proficient.” The expected interpretation is that “a high value indicates a large number of primary pupils are being taught in a language in which they are proficient thus making it easier for them to adapt to the school learning environment.”

This would not necessarily assume that it is only one language to be considered but all languages that the children are proficient in from their out of school environment. It is beyond this paper to discuss the assumption that children grow up with a single language until they visit school. Alidou (2006, p. 36) assumes child multilingualism in Africa as a natural situation. He sees in African *linguae francae*, that predates colonial times, a particularly high potential for extended use including education and inter-African communication. Indicators focusing on the idea of children with single home languages could eventually not reflect on education policies focusing on such languages.

However, the following review shows that a focus on one single language is almost a unavoidable when existing survey should be used and pooled. The data sources reviewed work almost exclusively with concepts of one language of instruction (or test) and a single home or first language of the child. Only for mother tongue some surveys record several languages.

In the data sources reviewed here children’s languages were operationalized in different ways, for some surveys even within one set of questionnaires. The concepts are used as items:

- First language / mother tongue:
 - The language(s) the child learned first, before going to school or “when being little”
 - The language(s) a child spoke before it started school
- Home language/ used languages in everyday life:
 - The language the child speaks (most of the time) at home
 - The language the child speaks (most of the time) at home with adults
 - The language(s) the mother / father (or respective guardian) speaks most with the child
 - The language the child speaks outside school

Kosonen (2017, 17 [based on Owen, 2015]) introduces the concept of a pupils strongest language. The proposal illustrates how complex items concerning languages can be. All data sources reviewed here fall short of such complex approaches.

A further important difference between data sources is that sometimes the concrete language(s) are recorded while otherwise simply the relation to a given language of instruction is recorded. The second option is most commonly used in student achievement studies since the analytical aim of those studies is not to report on different languages but to explain differences in achievement within an education system.

Frequent changes in PIRLS/TIMSS questionnaires from 2001 to 2011 illustrates the efforts by international studies to improve and optimize questions concerning languages while at the same time to keep the response burden low. The number of changes seems also an expression about ongoing uncertainty about how to best capture language issues for analysis. It seems that far more detail of data is collected than is used in monitoring or research.

PIRLS/TIMSS Questions on languages changed several times between survey. The following analysis draws therefore on different survey rounds to discuss certain aspects of the items used.

With the parallel use of the definitions “first language” and “home language” in one survey, it can be looked how the definitions relate to each other and whether interchangeable is possible.

The following is analysed:

- how frequent students speak the language of instruction at home?
- how consistent is the children’s report on first language with their parents report?
- how consistent is the children’s report on the home language with their parents report?
- how consistent is the children’s report on first and home language when reported both in the same questionnaire?

3.2.1 How frequent speak students the language of instruction at home?

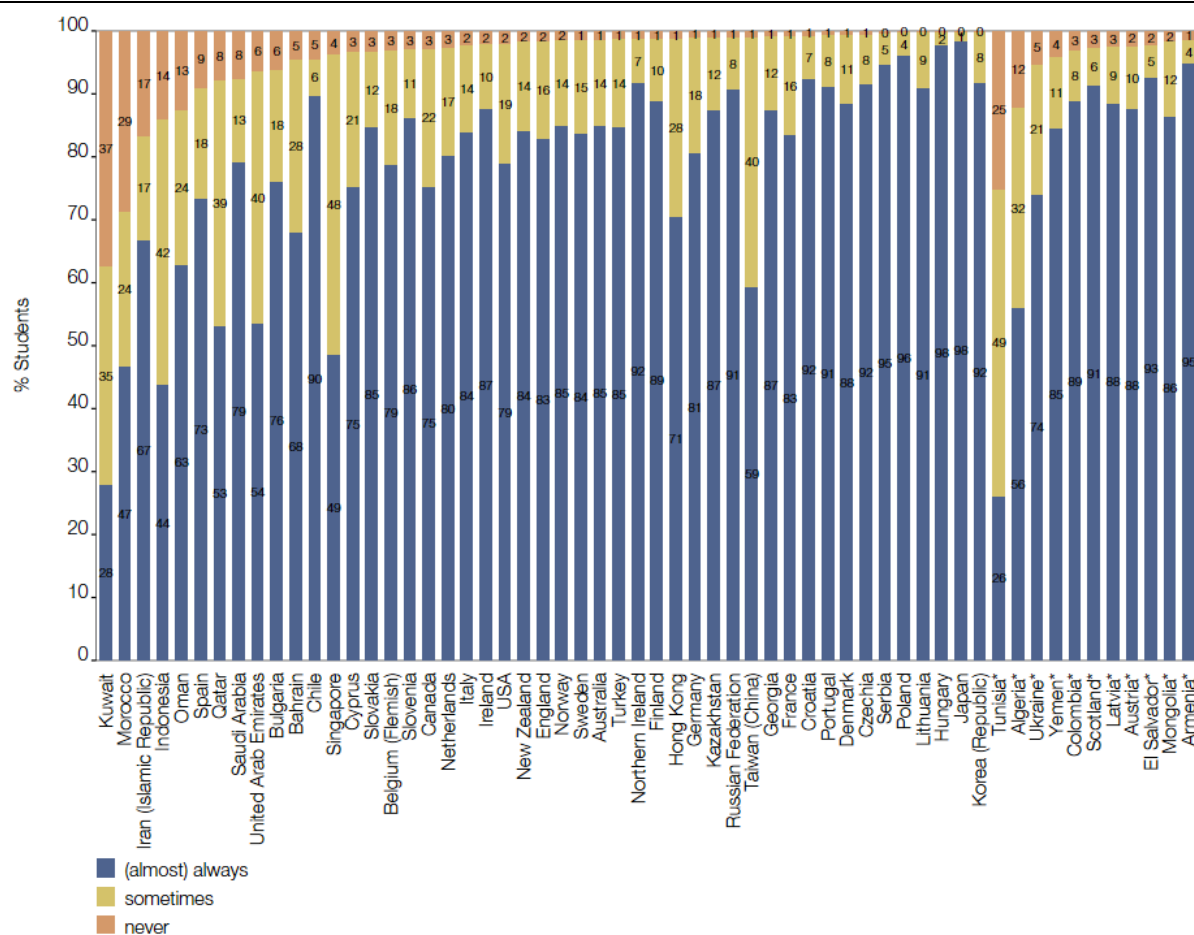
In TIMSS and PIRLS fourth grade students were asked “How often do you speak <language of test> at home?” Yet, the different rounds of TIMSS/PIRLS used different categories. TIMSS 2015 and TIMSS 2007 use four categories: always, almost always, sometimes, never. In TIMSS/PIRLS 2011 only three categories were used: always, sometimes language of test sometimes another language, never. (see Annex, e.g. Questionnaire example 4)

Figure 2 shows the results for the participating countries in TIMSS 2015 and for countries that participated in TIMSS 2011 but not in TIMSS 2015 (marked with *).

The group of students who report to never speak language of test at home is comparatively small. Just in Kuwait, Morocco, Indonesia and Oman (TIMSS 2015) and Tunisia and Algeria (PIRLS/TIMSS 2011) exceeds this share 10% of all students. Except for Iran (Islamic Republic), Kuwait and Morocco, the group who speaks sometimes but not always the language of test is bigger than the group that speaks never the language of instruction at home. In 8 out of 10 countries the group who speaks sometimes the language of instruction is more than three times bigger than the group that speaks it never. In PIRLS/TIMSS countries, students’ families not having the language of instruction as single language are rather bilingual than monolingual in another language.

If indicator 4.5.2 is to be gathered from different sources with different definitions of languages and frequencies of use, it is necessary to define “home language” for each source dichotomously. Therefore, clear cuts are needed to distinguish: language of instruction is home language or not. For TIMSS 2015 UNESCO (2017, p. 192) implicitly suggest a split at never or sometimes.

The relative small number of students in grade 4 who still never speaks language of instruction at home (not even with siblings) suggest to set the split not between never and sometimes. To split between almost always and sometimes would consider the language of instruction as home language only if spoken almost always rather than sometimes.



* 2011.

Source: TIMSS2015, TIMSS2011, PIRLS2011. Authors own calculations. See Annex X.

Figure 2: Students report on the use of test language at home.

Whether this split is also useful when just three categories are used, as in TIMSS/PIRLS 2011 and PIRLS 2006, should be questioned. If the middle category “sometimes” is contrasted to “always” and it indicates also parity between two languages (“sometimes language of instruction, sometimes another language”), it is likely that bilingual students would answer in “sometimes” even though language of instruction the most used home languages (See Questionnaire example 5, Questionnaire example 6). To judge this question results are compare for a group of countries that participated in studies with three and four answer categories. This is done by comparing results from TIMSS/PIRLS 2011 (three categories), TIMSS 2007 (four categories) and PIRLS 2006 (three categories). The differences show the impact the number of categories has. Especially the comparison between PIRLS 2006 and TIMSS 2007 shows this, since only year is between the surveys and the changes hardly reflect real change in all those populations. In 2007, when four categories were used, the percentage of students reporting to speak language of instruction always or almost is in all countries participating higher than in the years before and after when three categories were used. Between 2006 and 2007 most countries show a

difference of more than 10%-points. It can be concluded, that using three categories (never, sometimes, always) and to split between sometimes and always overestimates the share of children not speaking the language of instruction.

This analysis emphasises that data with similar questions but different answer categories lead to results with reduced comparability. Although a combination of different data sources may still provide in tendency correct results for a single point of time, reporting change should be done with great caution, when the items used are changing. For future reporting a recommendation on the categories to be used should be given.

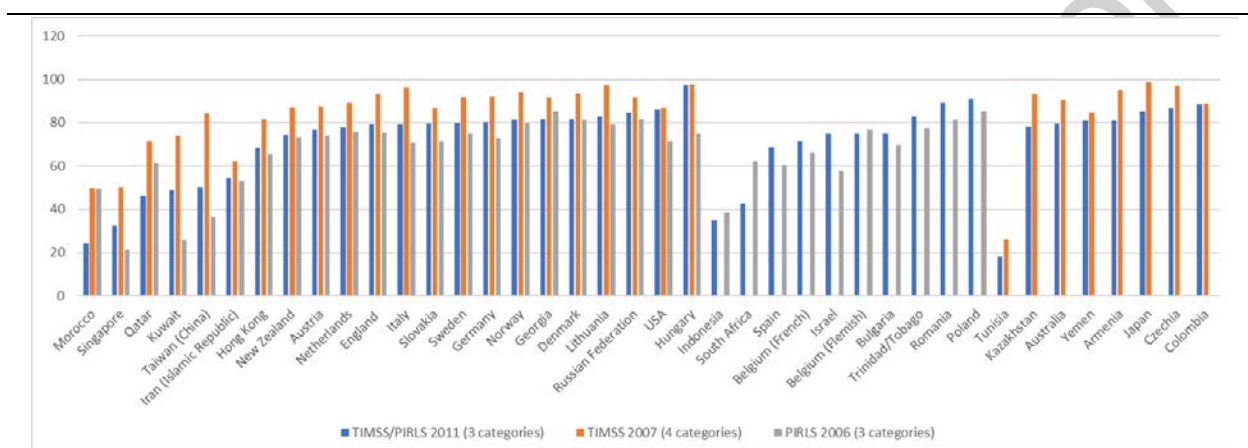


Figure 3: Comparison of two ways to ask for home language: Students who report to speak language of instruction always (or almost always)

3.2.2 Are statements according to students' first language reliable?

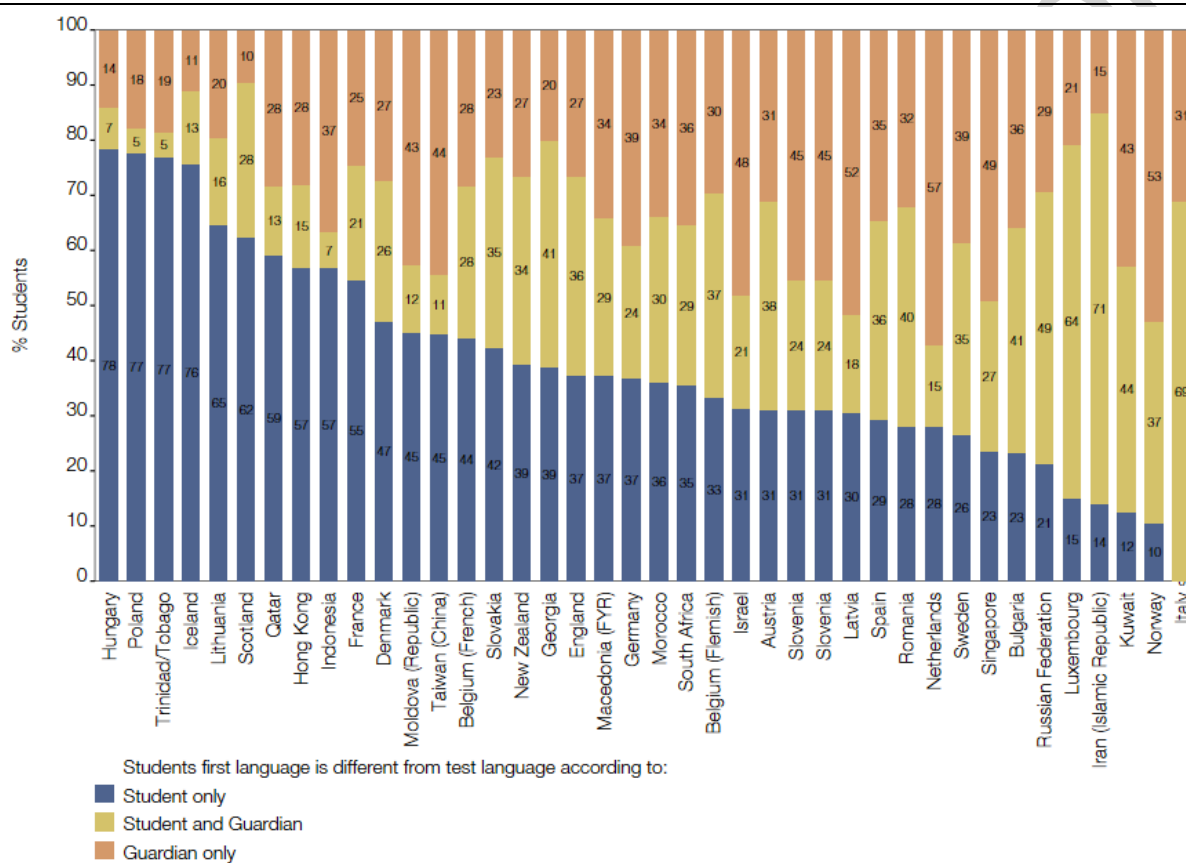
In PIRLS 2006 the student questionnaire was complemented by a home questionnaire administered to the parents/guardians of the students. This allows to check the reliability of the answers. Both questionnaire ask what language(s) the child spoke before it began school (See Annex Questionnaire example 6, Questionnaire example 2 and Questionnaire example 3). For children who learned several languages before school, multiple answers are possible. For indicator 4.5.2 the criterion of interest is, whether the language of test is the first language or one of them. The interest is not whether the child learned additionally a language different from the test language. Students were classified whether they learned the language of test as one of the first language(s) or not. The no category applies to cases when another language is answered with yes, but for the language of test no is answered. Figure 4 shows remarkable inconsistency between parents and their children. The figure focus on children, who according to at least one of the two questionnaires did not learn the language of test, meaning either the students themselves reported this or their parents or both.

With the focus on children not having learned the language of test¹⁰: For all countries except three, more than half of the cases are inconsistent. For 17 out of 40 countries less than a quarter of parents and students agree. The biggest group of students with inconsistent reporting are students that report

¹⁰ Reported either by the parent or by the child or by both.

having learned the language of test as one of several first languages, while the parents do not confirm this. But the number of students whose parents report that their child learned the language of test before school even though the children do not report this, is almost as high. That means, no clear pattern can be observed.

This analysis indicates that questions on first language or mother tongue are not providing reliable results and caution should be used especially when questionnaires are administered to primary aged children.



Note: 100% represent the children for who at least one questionnaire (home or student) indicates that they did not learn the language. Children with consistent reporting that they learned the language of instruction are not considered.

Source: PIRLS2006. Authors own calculations. See Annex X.

Figure 4: Students and guardians report on the first language of the student: Dis-/agreement between respondents

3.2.3 Are statements according to student's home language reliable?

TIMSS 2015 allows to check the reliability of the answers concerning the languages used at home. The students were asked how often they speak language of test at home (See Questionnaire example 6).



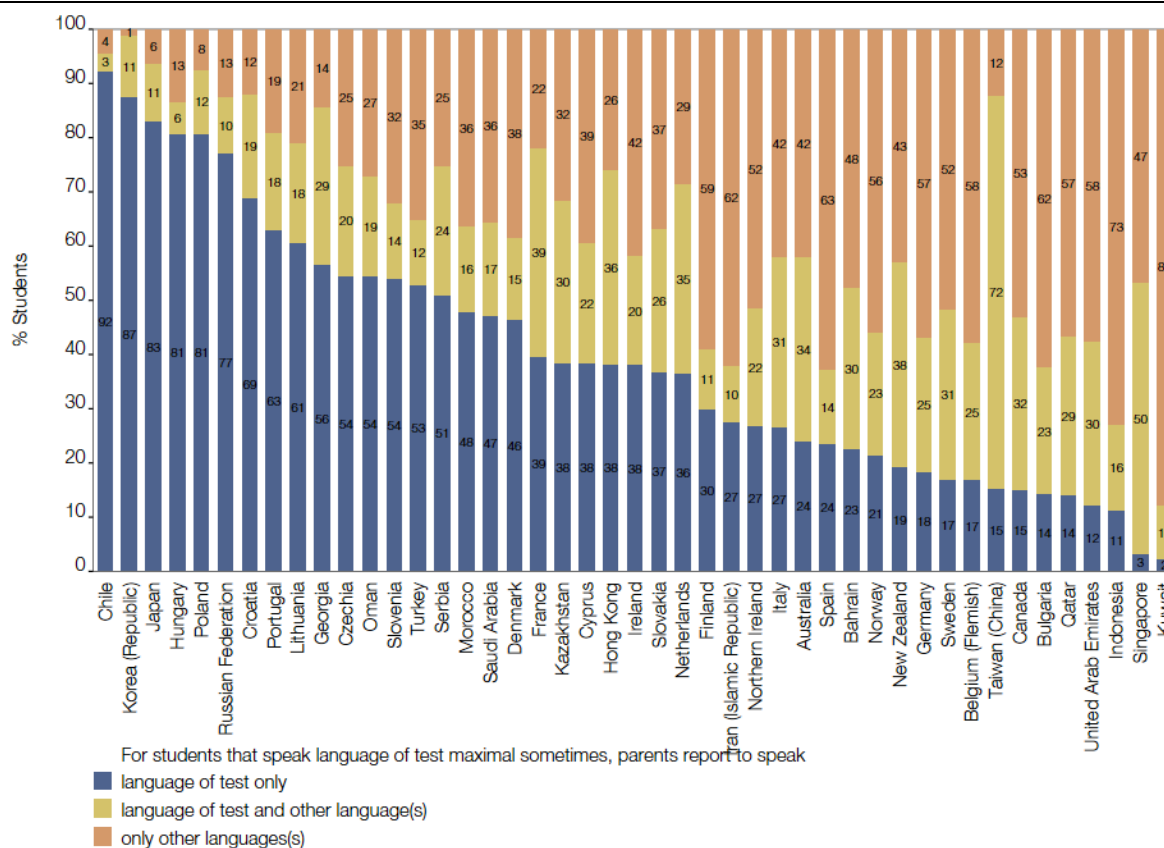
The parents/guardians were asked which language the father speaks most often at home and which language the mother speaks most often (see Questionnaire example 9).

Due to the number of bilingual couples, the one response by the student cannot simply be matched to the two parent's responses to check the validity. However, for students speaking the language of test maximal sometimes at home, it could be expected that at least one parent speaks another language. Figure 5 shows that for students who report to not speak the language of test at home (almost) always numerous inconsistencies can be observed with the response by their parents.

In more than one third of the countries there is relative strong consistency between students and guardians report on home language use. Here more than 70% of the guardians report, consistent with the child, that other languages are spoken. Yet, in one third of the countries the responses are surprisingly inconsistent. For more than half of the students who report to only speak sometimes or never the language of instruction, the parents report language of instruction as only language used by the adults.

This analysis indicates that also questions on home language are only reduced reliable for fourth grade students.

DRAFT for discussion



Source: TIMSS 2015. Authors own calculations. See Annex B, Table 5.

Figure 5: Guardians report on the home language of students who say they speak language of test at home never or sometimes

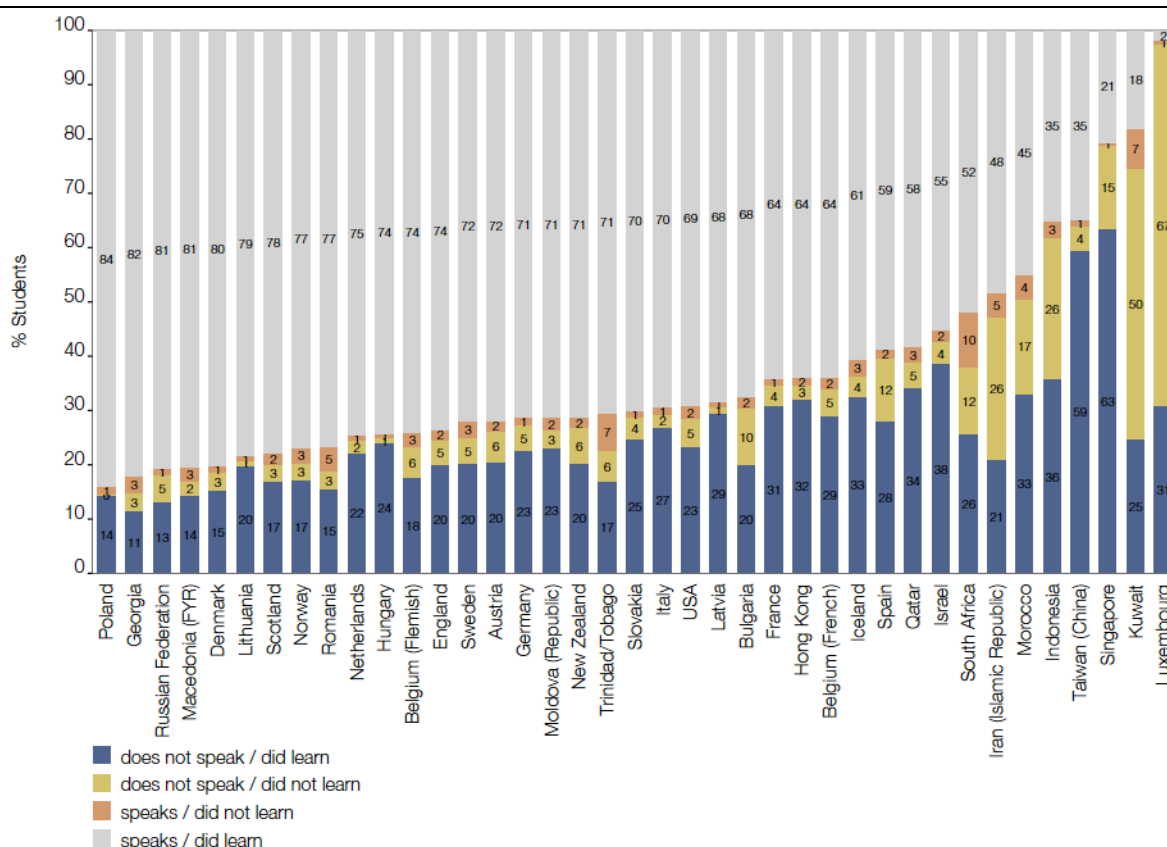
3.2.4 How consistent is reporting on first and home language when students report on both?

In PIRLS 2006 students were asked two questions concerning language. They were asked which language(s) they learned before they started school and how frequent they speak language of instruction at home (three categories). It could be expected that students reporting the language of instruction as home language (they speak it always) also report that they learned the home language before school. Figure 6 shows the distribution of students by the four different situations:

- students learned language of instruction before school and speak it always at home,
- students learned language of instruction before school and but speak it only sometimes at home,
- students did not learn language of instruction before school and but speak it always at home,
- students did not learn language of instruction before school and speak it only sometimes or never.

The figure shows for TIMSS 2015 countries again, that only a very small proportion of children report that language of instruction was not learned before school. Therefore there is a big group in most countries of students that report to have learned language of instruction without speaking it. The

group of students that speak the language always but report to not have learned it is comparable small.



y

Source: PIRLS2006. Authors own calculations. See Annex X.

Figure 6: Students report on the use of language of instruction at home and on having language of instruction as first language

3.2.5 Details on specific languages

Most data sources reviewed collect some data on which specific language the child's uses or learned besides language of instruction. Only PASEC and SECMEQ do not collect detail on which language the child can speak except for the language of instruction. TIMSS/PIRLS asks only for the language learned first. The reliability of the data on first language is questionable (see above) and consistency with home language (as used by the other surveys) is not given. As a consequence information on first language from TIMSS/PIRLS should not be mixed with information on specific home languages from other studies. As a consequence, data from different sources on specific languages should not be pooled for SDG reporting. Analysis on specific languages should always be done in the context of individual data sources.

Yet, when advising countries in their data collection, it seems necessarily to understand the details of the countries situation. Countries designing national questionnaires would be better advised to take



items from EGRA, LLECE, PISA or MICS as model (See Annex): Questionnaire example 19, Questionnaire example 22, Questionnaire example 23, Questionnaire example 24 or Questionnaire example 25.

However, recording languages correctly is not a straightforward task. Countries need to be cautious when deciding on the list of languages provided in the task to decide on the child's language is not straightforward (Kosonen, 2017, p. 17).

3.2.6 Operationalization of student's language

The section discussed so far, some general aspects related to items on students language. Since PIRLS/TIMSS data are public available, the analysis was limited to questions that could be observed using those data. In the following the items of all reviewed data sources are introduced.

3.2.6.1 IEA-Studies: PIRLS, TIMSS, PreTIMSS

Over the different rounds of TIMSS and PIRLS studies numerous items on student languages were used in student and home questionnaires. An overview of most item can be found in Annex A: 5.10. The most recent available survey, TIMSS 2015, included items in the home questionnaire for parents/guardians and in the student questionnaire:

- Parents: First language: language spoken before school
- Parents: Languages used by father/mother when talking to child at home
- Parents: How often the child speaks language of test at home?
- Students: How often do you speak <language of test> at home?

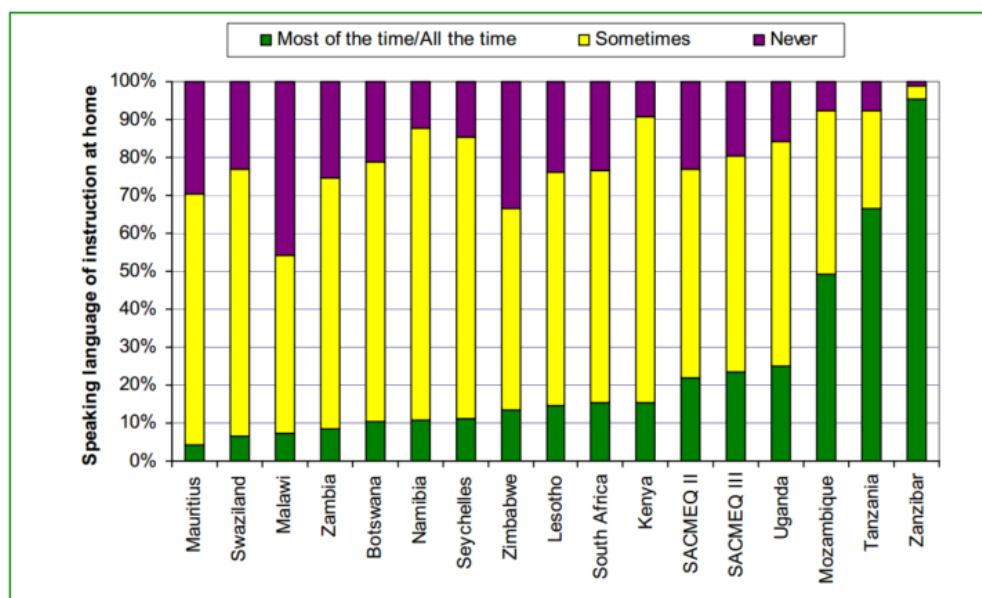
The above analysis gave some impression how hard it is to interpret different items on languages from the same survey. Therefore, it seems, that even when different items were included in the questionnaire, the international reporting of data is focused on one perspective on student languages. In the case of TIMSS/PIRLS this is home language as reported by students, even when other items are available, as in PIRLS 2006 (Mullis et al., 2007, p. 288). This is also due to the fact, that some countries do not administer the home questionnaire.

3.2.6.2 SACMEQ

In the different rounds of SACMEQ (I to III) the items on the language of students were similar to the core items in PIRLS/TIMSS on the home language of students, but it is not focused on the home of students but general on "outside school". Thus the data reflect on everyday life and not narrowly on the home of students. Students were asked how often they spoke the language of instruction outside school. SACMEQ also used different versions of the item in different rounds of SECMEQ. While SECMEQ II used three categories for the item, SACMEQ III used four categories, including most of the time and all the time.

For the purposes of reporting Hungi (2011, p. 13) identifies students speaking the language of instruction using a narrow definition. He considers students who reported that they spoke the language 'sometimes', 'most of the time', or 'all the time', as speaking the language of instruction at home. Figure 7 illustrates the importance of identifying the correct cut between categories. When the

way the authors of the GEM 2017 interpret TIMSS data (UNESCO, 2017, p. 192) would be applied to SACMEQ data, this would lead to a very different interpretation as by Hungi (2011, p. 13).



Source: (Hungu, 2011, p. 13)

Figure 7: Categories of pupils who spoke the language of instruction at home (SACMEQ III)

3.2.6.3 PASEC

For PASEC only questionnaires for 2014 could be reviewed¹¹. The items on the language of students were similar to the core items in PIRLS/TIMSS. Students were asked how frequent they speak <official language of instruction > at their home. For almost all countries here French was substituted as official language of instruction, when the latter is not French (as is the case of Kirundi in Burundi at the beginning of the primary cycle, and English in Anglophone Cameroon both at the beginning and the end of the primary cycle), questionnaires were adapted to the country's linguistic context (Programme d'Analyse des Systèmes Educatifs de la CONFEMEN, 2015, p. 20). In PASEC 2014 three answer categories were given: "I speak always French", "I sometimes speak French and sometimes another language" and "I never speak French".

In previous round of PASEC obviously more details on languages of students was available in the data. Fehrler and Michaelowa (2009, p. 10) reports very detailed on different language groups related to achievement: Agni (CI, N=113); Bamileke (CM, N=265); Baoule (CI, N=337); Bassa (CM, N=130); Bete (CI, N=130); Bissa (BF, N=68); Dioula (BF, CI, SN, N=370); Douala (CM, N=42); Ewondo (CM, N=553); French (all, N=3454); Fulfulde (CM, N=35); Gourmantchi (BF, N=83); Maka (CM, N=53); Malagasym (MD, N=64); Mandingue, (SN, N=75); More (BF, N=975); Other (all, N=665); Peul (BF, N=28); Guere (CI, N=106);

¹¹ Which the PASEC centre made promptly available, which the author wants to thank for.



Gourounsi (BF, N=54); Pulaar (SN, N=260); Senoufo (CI, N=86); Serere (SN, N=242); Soninke (SN, N=25); Wolof (SN, N=842)

The very small sample sizes related to the different should be considered problematic for reporting and monitoring of linguistic subgroups. Despite scientific interest, global monitoring efforts should avoid to gather data to such detail.

3.2.6.4 *The Latin-American Laboratory for Assessment of the Quality of Education (LLECE)*

LLECE includes different items on the language of students. The home language is obtained from students and the first learned language from the parents.

In the student questionnaire students are asked which language they speak at home. (See Questionnaire example 19 in Annex) Students are given a choice of languages with the limit to check only one language. In this way students are forced to decide on one language which leads to a dichotomous answer when calculating a variable indicating whether language of instruction is the home language.

If the dichotomous LLECE data on speaking language of instruction should be matched with TIMSS, SACMEC or PASEC, it seems that recoding TIMSS, SACMEC or PASEC in a way that speaking “always/almost always” is interpreted as speaking the language of instruction would give a better match as if “sometimes” is also interpreted as speaking language of instruction.

In the home questionnaire parents/guardians were asked which language the child learned first (See Questionnaire example 20 in Annex). Here also only one choice is possible.

When LLECE ask for the languages, it classifies languages already in the questionnaire as Spanish/Portugese, Foreign language and indigenous languages.

3.2.6.5 *Early Grade Reading Assessment (EGRA)*

The implementation of EGRA differs between countries where it is implemented. For this review a version of the questionnaire is available from the School Management Effectiveness studies of USAid. It is not clear to what extend the questionnaires are standardized.

As in LLECE, EGRA does not assume a default language, but in the questionnaire students are asked which language they speak at home. Two fix options are given (English and one other), but an open-ended option is available. Students must decide on one language as home language.

3.2.6.6 *MICS*

With MICS6 information on the language of the children will be gathered from the children and their guardians directly. The child is interviewed together with the mother, where the face to face situation with probing the language might lead to good reliability of the answers. The children will be asked which language they speak most of the time at home. They have to decide on one language, which allows good match with the question on the language of instruction. The item used is comparable with the corresponding LLECE item. The list of language options is identical to the list given when asking for the language of instruction (see Questionnaire example 25 in Annex).



3.2.6.7 National assessments

Many countries introduced in recent years national assessments, that could be considered as data source for international reporting. For this paper a few nationally used student questionnaires were reviewed. Many of the questionnaires are obviously influenced by the countries prior participation in international and they use similar items as the discussed international assessments.

The concept of home language is used in several national assessments. Examples are as follows: Germany uses as concept of “family language” asking for frequency of use of German at home with three categories (Stanat & al., 462). Austria uses the term “every day language” for home language and asks also for the frequency of use of German at home in three categories (Questionnaire example 27). The NAEP assessment in the United States follows another approach. The student questionnaire does not record the language of students, but for students classified as “English Language Learner” a special questionnaire is conducted (as interview). Here the “primary language” is recorded as Spanish or other.¹²

3.2.6.8 EMIS Systems

Education census and other administrative routine data collections (commonly referred as “EMIS”) could also serve as national data sources to monitor instruction in the home/first language. Most EMIS data collections are done at the school or class level in form of school aggregated data. In such systems, the collection of student level characteristics is currently not common, but it exists. The EMIS system of Scotland collects aggregated data from schools on the language of students. But it appears, that this is possible because the system relies on school management systems that host student level data (ScotXed, 2017). Also some German states collect language of students as aggregated data (Kemper, 2017).

Other countries build up student registries which include student level information. Mellander (2016) sees register data as a fundamentally Nordic (North European) phenomenon. Yet, other countries start building up similar systems. Mellander reports that data on the mother tongue of pupils are a common feature of Nordic registers. Unfortunately, this could not be confirmed by documents of the responsible statistical offices.

Beginning in 2003, Austria build up a student register following the law on documentation of education (“Bildungsdokumentationsgesetzes”). Each student in the school system is registered with a unique number linked to the student’s social security number (Statistik Austria, 2014). For each student the language used in every day (“Alltagssprache”) life is recorded by the schools when students enter the school. Since the data can be linked over the years, this information obtained in the first grade of each level is also available when students move to higher grades. Data are used for national education monitoring (e.g. Bruneforth et al. 2016, p. 43, 121). Other examples of collection of student individual data in registers can be found in half of the German states (Kemper, 2017), England (Department for Education, 2017) and Ireland (Department of Education and Skills).

A special form of collection of national student level data are student/parent surveys not in the context of assessment studies. The Australian state of Victoria used a student surveys with the school census.

¹² https://nces.ed.gov/nationsreportcard/pdf/bgg/sch-sdlep/2013_BQ_ELL.pdf



Parents are obliged to report on their family's characteristics (e.g. parental education, occupation, country of birth) including home language. (Department of Education and Early Childhood Development, 2013). In similar way collects the education system in Flanders (Belgium) data from parents (Flemish Department of Education and Training, 2015, p. 29). In England parents are obliged to participate in the school census and report on language (Department for Education, 2017).

As for other data sources, the national education systems use a variation of definitions for student languages. Most systems use some definition related to home or family language. In some cases, this is not clearly distinct from mother tongue.

The following examples show the variation of definitions.

In the English school census language is defined as follows: the pupil has been exposed to a language other than English during early development and continues to be exposed to this language in the home or in the community. If this condition is observed, the specific language of the student is reported, if possible. (Department for Education, 2017, p. 64).

Kemper (2017, p. 155) shows the variation of wordings for the 11 out of 16 German states that cover student language:

- Non-German language in family (independent whether student is able to speak German)
- Language mainly spoken in family
- Main non-German language in family
- Languages spoken at home (in addition to German)

This list indicates similar uncertainty in the definitions as in assessment data. For some states the focus is on whether any other languages are spoken in the family, while in others the emphasis is on main language.

A mother tongue or first language as definition is not common. Ireland uses mother tongue as a term, but it is used as a synonym to home language. The question in the Irish student register is "Is one of the pupil's mother tongues (i.e. language spoken at home) Irish or English? (Yes/No). The German state of Bavaria plans to move from home language to the use of mother tongue, which is seen as problematic (see Kemper, 2017).

3.2.6.9 *Census and national household surveys*

Language use is one of the traditionally defined demographic and social characteristics investigated in population censuses (United Nations, 2008, p. 238). According to United Nations (2008, p. 138), there are three types of language data that can be collected in censuses, namely: (a) Mother tongue, defined as the language usually spoken in the individual's home in his or her early childhood; (b) Usual language, defined as the language currently spoken, or most often spoken, by the individual in his or her present home; (c) Ability to speak one or more designated languages. The UN Recommendations do not give priority to any of the options but emphasise that international comparability is not a major factor in determining the form of the data to be collected on this topic. Yet, for the purpose of indicator 4.5.2 only data based on the definitions a) and b) could be a basis for calculations.

For national household surveys it can be assumed that the UN recommendations on census and / or examples from international household surveys are quite similar to common national surveys covering languages of individuals.

4 Sampling issues

A good understanding of processes and definitions used for sampling is important in order to judge the feasibility to generalize from survey data to the whole country and to the primary level of education as a whole.

Three aspects need to be considered:

- Study design: Which populations are represented by the survey.
 - Which grades are tested
 - Which populations are included
- Sample Implementation: Is the population coverage unbiased for the purpose of data use
- Sample sizes.

This section does not aim to evaluate the quality of the sample of the studied surveys. Yet, his paper is about pooling data from different surveys for a purpose that is different from the original purpose of the surveys. Therefore it must be asked whether there are issues in the sample design and implementation that are especially relevant when reporting on languages of students and instruction.

4.1.1 Focus on grade or on level of education

With the notable exception of MICS, all data from other discussed survey are designed to represent students of a single grade (or two) of primary education. This is summarized in the following table:

Survey	Lower grades	Higher grades
TIMSS/PIRLS		grade 4
SACMEQ		grade 6
PASEC	grade 2	grade 5
LLECE	grade 3	grade 6
EGRA	grade 2,3, sometimes 1	in exceptions 4
MICS	all grades, children are sampled independent of grade	

Representative national assessments tend to be conducted at higher grades (4 to 6) to summarize the learning outcomes of the primary level of education. The UIS data base on national assessments¹³ lists assessment programmes from 31 countries, mainly developing countries, of which 9 countries conduct test in grades 1 to 3 and the other only in higher grades of primary education. OECD (2013, 4.A2.5a) lists data on compulsory national assessments in primary education for 16 education systems, of which 7 conduct assessments in grades 1 to 3.

Using data on language of instruction for single grades poses serious challenges when the goal is to generalize from the given grade to the complete primary education. While it is reasonable to interpret outcome data as summarizing the level of education up to the tested grade, descriptive statistics on class room processes cannot be generalized from higher to lower grades. At least this cannot be done for systems with multilingual education policies (MLE) policies.

¹³ <http://nada.uis.unesco.org/nada/en/index.php/catalogue/>, accessed 30.10.2017

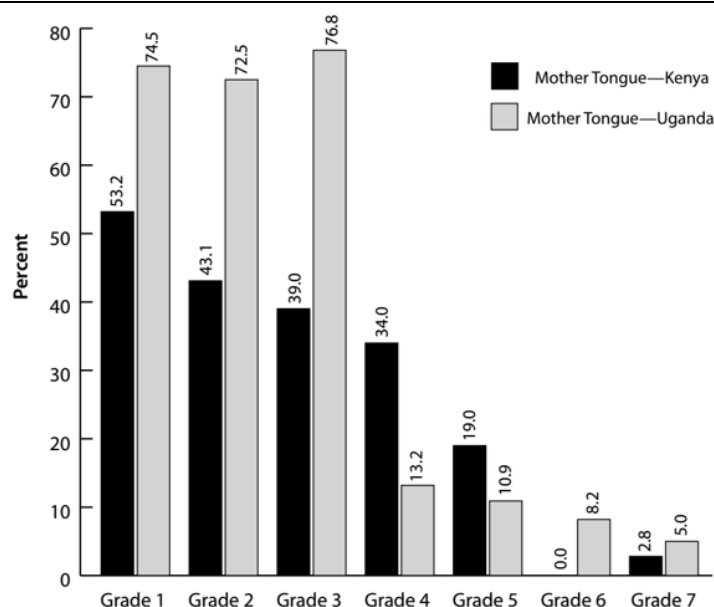


The proposed indicator 4.5.2 is intended to reflect on students in primary education, which must be interpreted as primary education as a whole. For systems with MLE this raises a number of questions, especially when ONE main language of instruction is defining for the indicator. Is the perspective cross sectional? That would mean the language of instruction is identified for students dependent on the grade they are actual in. Consequently, for each grade the students are classified according to the language of instruction in the given year and the indicator would be aggregated. It must be noted, for many data sources information is only available for one given grade. Alternatively, a level-wide perspective would ask for the dominant language of instruction over the span of all primary grades. Students currently in grade 4 or 5 who benefited from 3 years instruction in their home language could be classified as having their home language as main language of instruction in primary education.

Clear is, that the language of instruction observed at the end of primary (grades 4 to 6) would not be sufficient to reflect on the primary education as a whole. This limits the use of student assessments to gather information on 4.5.2. National and international assessments are mostly conducted towards the end of primary education and thus likely after the transition of languages in systems with MLE is completed. The same applies to national assessments. In the case of Ghana, the assessment was explicitly moved from grade 3 to grade 4 to better align with Ghana's current language-of-instruction policy (Ministry of Education Ghana, 2016, p. 1). Thus, using Ghana's assessment data would show a low percentage of use of home language of instruction, partially as artefact of the timing of the assessment.

South Africa's participation in PIRLS/PrePIRLS illustrates clearly the impact of the grade observed on results for 4.5.2. South Africa tested in 2011 for the PIRLS main study students in grade 5 and for PrePIRLS students in grade 4. For the lower grade 11 languages of instruction were considered. With grade 5 there are just two main languages of instruction remaining, English and Afrikaans. Calculating the percentage of students who have a home language different from the language of instruction (in PIRLS: language of test) for grade 4 and grade 5 leads to substantially different results. In primary grade 4 there are 65% of students whose first or home language is the language of instruction. In primary grade 5 this number is down to 42% of students whose first or home language is the language of instruction.

Gove and Wetterberg (2011) shows a similar situation for Kenya and Uganda. In both countries there is a transition from lower to upper primary in between different language of instruction used.



Source: Gove & Wetterberg, 2011, p. 163.

Figure 8: Percentage use of mother tongue as language of instruction in schools in Kenya and Uganda

4.1.2 Out of school children

Indicator 4.5.2 is formulated to cover primary education and not to inform on all school aged children. However, it should be noted that being in or outside school correlates as much with language aspects as language aspects correlate with achievement. Pinnock (2009, p. 8) indicates that language groups are commonly excluded from education. Pinnock (2009, p. 9) reports that 58 % of out-of-school children live in “highly linguistic fractionized” countries. Students belonging to language groups for which education is not accessible in their language tend to be more frequent out of school. Using samples of students within school to address language issues might misrepresent the lack of access to education in the language of instruction in countries with high rates of out-of-school children. Smits and Gündüz-Hoşgör (2003) based on DHS-1998 found for Turkey that “over half of the (mostly rural) girls who did not go to primary education had a mother who was not able to speak Turkish”.

This problem is even more relevant for reporting 4.5.2 when focusing on grade based assessments. Students who enter grade 4 fall in the target population. Developing countries can have very high dropout rates until grades 4 or 6. If the share of students having instruction in their home language is monitored at higher grades, a substantial share of the vulnerable language groups are likely to have dropped out. In Ethiopia, Madagascar, Uganda, Chad, Mozambique, Burundi, Cambodia more than 33% of grade 1 students drop out before reaching grade 4 (<http://data.uis.unesco.org/>, reference years 2012 to 2014). In those countries indicators based on grade 4 assessments are unlikely to be generalizable to all primary education. In thirty out of 120 countries with data available more than 15% of children do not reach grade 4 and would not be covered in late assessments. Salchegger and



Suchań (2017) shows for PISA data that already relatively low out of school rates can distort descriptive data substantially. It is likely that comparing countries with respect to the language of instruction is problematic, if the data sources cover different grades.

4.1.3 Target populations and exclusions

Sample design is important to properly monitor equity. UNESCO (2017, p. xvi) reports that “inequality is underestimated, as survey design may exclude up to 250 million vulnerable people worldwide, while a further 100 million, such as slum dwellers, may be under-represented.” Therefore sample survey must provide good documentation on their samples to be reliable sources for monitoring at national, regional or global level. As mentioned before, large scale assessments are not designed with the target to provide descriptive statistics on variables not related to the tested domain. For example, Austria excludes all special schools from its national assessment at grade 4 because the educational standards to be monitored do not apply to those schools (Breit, Bruneforth, & Schreiner, 2016). This design based exclusion affects about 1% of students that speak the language of instruction at home, but 5% of students with another home language than the language of instruction (Bruneforth et al., 2016, p. 97). The following sections provide some information on sampling issues related to the reporting of home and instruction language. The review is by far not complete. The most details are taken from the TIMSS/PIRLS studies. That does not indicate quality issues with the study but reflects the good quality of the accessible sampling information.

4.1.4 IEA-Studies: PIRLS, TIMSS, PrePIRLS

The target population for TIMSS (and PIRLS studies) is “All students enrolled in the grade that represents four years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 9.5 years” which is typically grade 4 in a country.

IEA studies document not only international definitions for sampling but provide a detailed documentation of deviations. The following section is mainly based on LaRoche and Foy (2016).

TIMSS/PIRLS is designed to summarize student achievement across the entire target grade, and national target populations should aim for comprehensive coverage. But this is not achievable and due to some cases, political, organizational, or operational factors deviate countries from the intended coverage. Reduced population coverage is not unbiased. Exclusions of schools and students in national samples in TIMSS represent typically less than 5% of the international target population, even though in a small number of countries it exceeds this limit. This exclusion rate is deemed small enough to ensure good and comparable measurement of TIMSS key statistics, mainly achievement. Yet, some descriptive statistics may be biased by exclusion.¹⁴

Countries are permitted to exclude schools on the following grounds when they consider it necessary:

- Inaccessibility due to their geographically remote location
- Extremely small size (e.g., four or fewer students in the target grade)

¹⁴ A simple example is related to the exclusion of very small schools. It has little impact on achievement data. Yet, statistics on average school size would deviate substantially from statistics with full school coverage.



- Offering a grade structure, or curriculum, radically different from the mainstream educational system
- Providing instruction solely to students in the student-level exclusion categories listed below (e.g., catering only to special needs students)

The TIMSS/PIRLS student-level exclusion rules are:

- Students with functional disabilities — Students who have physical disabilities such that they cannot perform in the TIMSS testing situation. Students with functional disabilities who are able to perform should be included in the testing.
- Students with intellectual disabilities — Students who are considered, in the professional opinion of the school principal or by other qualified staff members, to have intellectual disabilities or who have been tested as such.
- Non-native language speakers — Students who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier in the test situation. Typically, a student who has received less than one year of instruction in the language(s) of the test should be excluded.

The list makes clear that coverage of student assessment surveys is not unbiased towards inclusion of special language groups. It can be assumed that this is the case for most other surveys. In TIMSS 2015 non-native language speakers were excluded in all countries except: Egypt, England, Hong Kong, Indonesia, Iran, Jordan, Kazakhstan, Kuwait, Malaysia, Morocco, Northern Ireland, Singapore, South Africa, Thailand.

School exclusions in TIMSS were documented for each country. Most of them are related to issues of the language of test, as the following:

- United Arab Emirates: language other than English or Arabic
- Slovenia: Italian schools
- Slovak Republic: schools where language of instruction is not Slovak or Hungarian
- Serbia: language other than Serbian
- Qatar: instruction not in English or Arabic
- Portugal: minority language schools
- Poland: language of instruction other than Polish
- International schools: Canada, England, Spain, Sweden
- French schools (in Canada, Newfoundland),
- Bahrain students taught in French, and students taught in Japanese. NOTE: could be adjusted
- Cyprus: French language
- Geographically inaccessible schools: Chile, Chinese Taipei, Iran, Turkey, United Arab Emirates
- Czech Republic: Polish language schools
- Finland: instructional language other than Finnish or Sweden
- Hong Kong: schools teaching in Japanese
- Hungary: students taught in foreign language
- Israel: English or French schools
- Italy: Slovenian language schools, Ladin language schools, and German language schools
- Kazakhstan: Uzbek, Uighur, Tadjik only schools



- Kuwait: minority language schools
- Lithuania: language of instruction other than Lithuanian, Russian, or Polish
- Malta: Language schools.
- New Zealand: Maori-medium Level 1 immersion schools
- Norway: Sami language schools

Some school exclusions might not be directly linked to language, but may correlate high with it.

With school exclusions are documented in student tracking forms, but the information is not recorded in international data. The amount is typically small in respect to reporting of achievement, e.g. in Austria the exclusion due to language is 0.9% (Suchań & Schreiner, 2012, p. 74).

The target for sampling precision and sample size for TIMSS is set with a view on the achievement data. The students are sampled in a two stage cluster sampling with a sample of complete classes within schools selected proportionally to their size. The standard error for the mean achievement should not be bigger than 0,035 standard deviations. For most countries, the TIMSS precision requirements are met with a school sample of 150 schools and a student sample of 4,000 students for each target grade (LaRoche & Foy, 2016). When judging the sample size the cluster effect needs to be taken into account. The effective sample size is much smaller and would be equivalent to about 400 students in a random sample of students.

The sampling errors concerning the home language of students ranges between 0.2 and 2.3 %-points for the combined group of students speaking the language never or sometimes, with a median error across countries of 1 %-points.¹⁵ Sampling precision is thus not an issue for this data source.

4.1.5 SACMEQ

The sampling process of SACMEC is similar to that of PIRLS and TIMSS, except that not complete classes were drawn but 20 students per school. The target population for SACMEQ 3 was “All pupils at Standard 6 level [...] who were attending registered mainstream primary schools.” (example: Kenya: Wasanga, Ogle, & Wambua, 2012, p. 13). The use of the word “mainstream” in the definition indicated that special schools for the handicapped should be excluded from the SACMEQ III. In Kenya also small schools were excluded– based on the definition of having less than either 15 or 20 pupils in the desired target population. They were mostly located in very isolated areas.

The sample size in SACMEQ is similar to the of TIMSS PIRLS. Kenya sampled for SACMEQ 3.436 students in 193 schools. The effective sample size being 385 students. Also for SACMEQ data sampling precision is not an issue.

4.1.6 PASEC

The sampling process of PASEC is similar to PIRLS and TIMSS, except that within schools not complete classes are drawn but 20 students from all grade 6 classes and 10 students from all grade 2 classes. The target population are students in grade 3 and 6.

The sample size in PASEC is about 180 school with 3600 students in grade 6 and 900 students in 90 schools in grade 2. The schools selected for grade 2 are a subset of the grade 6 schools. A detailed

¹⁵ Calculated using Jackknife methods with the R-package BIFIEsurvey.



documentation of exclusions was not available for this review. Also for SACMEQ data sampling precision is not an issue.

4.1.7 The Latin-American Laboratory for Assessment of the Quality of Education (LLECE)

The sampling process of LLECE, as applied in the second study, SERCE, is similar to PIRLS and TIMSS. The target population are students in the 3rd and 6th grade. Not part of the target population are students enrolled in special education institutions, in non-formal alternative education programs and, for the 3rd grade, in schools catering for children whose mother tongue is indigenous. The latter is, because they do not yet master the language to test sufficiently. For the 6th grade, it is considered that all children participating in the test master the test language (Spanish/Portuguese). Excluded at the student level are students with special educational needs (LLECE, 2010, p. 13). Further can countries exclude small schools (less than 6 students enrolled in the grade) and 3rd grade students in bilingual schools with indigenous language (LLECE, 2010, p. 15). At the student level are students excluded whose first language is different from Spanish or Portuguese and who had less than two years education in the language. Exclusion of small schools sum up to 0.5% to 5.4% and 5.6% in (Colombia and Cuba). These exclusions are mainly schools in rural areas, as a consequence a quarter of participating countries excludes more than 10% of rural schools.

The sample size in SERCE is a minimum of 150 schools and results to a sample size of 4000 to 5500 students in grade 6 and 3500 to 5400 students in grade 3 (LLECE, 2010, p. 22). Sampling precision is not an issue.

4.1.8 EGRA

The goal of the sample in EGRA are national representative sample of the grade to be studied with a precision to provide information at the system level. The grades studied in EGRA are typically lower grades, grade 1 to 3, with exceptions 4 (RTI International, 2009, p. 60). EGRA is the only survey of schools covering the first two grades. Until data from MICS 6 become available, this is a unique source. EGRA describes itself as an attempt to respond to the demand for simple, effective, and low-cost measures of student learning outcomes (RTI International, 2009, p. 2). This is reflected in a less strict approach to sampling. It seems national samples are not quality assured and no common guidelines are enforced for sampling issues. The samples are drawn using a two or three stage cluster sampling (regions, schools, students). Yet, some samples have elements of judgement samples, such as EGRA in Kenya and Uganda (Gove & Wetterberg, 2011, p. 154) or are optimised for treatment studies and other research needs.

EGRA provides to countries a guideline to determine the sample size. The needed sample size is suggested to be sufficient to significantly distinguish the achievement differences between two grades. They suggest as orientation: "It is reasonable to ask, for example, that the confidence intervals for different grades not overlap each other." (RTI International, 2009, p. 61). The standard error, half the confidence interval, is thus one quarter of one year of learning. Drastically spoken, the sampling error is so wide that it is just possible to determine whether the learning gain in one year of schooling



is significantly different from zero¹⁶. Compared to the target sampling error of PIRLS (0,035 of a standard deviation in the achievement in one grade) is the target for EGRA much wider: 0.12 of the standard deviation of the achievement of the pool students in grade 1 to 4. As a consequence, EGRA defines sample sizes in clustered samples (cluster size 12) of about 410 to 600 students as sufficient (RTI International, 2009, p. 67). For practical implication a sample of just 400 students per grade drawn from just 20 schools is deemed sufficient (RTI International, 2009, p. 39). Yet, these guidelines seem to formulate minimum targets for samples. A number of implementation of EGRA survey exceed the here mentioned sample sizes per grade by far (RTI International, 2015, p. 24).

For this paper it must be reminded, the purpose here is not to judge the EGRA sample relative to the purpose of EGRA (monitoring early grade reading), but with respect to observe the match of home and instruction language for students. Instruction language is a variable that is highly clustered in schools and also home languages are clustered. That given, surveys with just 20 schools in the sample seem not sufficient to generalize to a national level for monitoring purpose or to observe change.

When EGRA should be used as source to monitor 4.5.2 this should be limited to country data based on bigger samples. If countries are not covered by other data sources, results from smaller EGRA samples could be used to inform regional or global estimates. But country level reporting on progress in SDGs should not be based on insufficient sized samples.

4.1.9 MICS

The MICS surveys aim for nationally representative samples from all households, not just those households that have school age children. The households with primary age children are a subset of all sampled and interviewed households. Yet, with proper weighting the data can be representative for the population of primary school age children or children in school.

The items on children's language of instruction and home language will be used in the newly introduced questionnaire for children aged 5 to 17. This questionnaire will be completed for one randomly selected child in that age group in each household.

For the purpose of reporting on students languages, will MICS6 sample of primary aged children be big enough to calculate reliable? In order to judge this question data from MICS 5 were analysed and the coming sampling routine for the child questionnaire simulated.

The number of sampled households in MICS5 survey ranges from 55,000 households in Bangladesh and 33,100 the Dominican Republic to 4,000 to 6,000 in Turkmenistan, Kosovo and Sao Tome and Principe. The median sample size is 14,000 households.

Table 1 provides an overview of the sample sizes in MICS5 with respect to all children aged 5 to 15, primary aged children and children in grade 1 and 2 only. Even when sub-setting the MICS samples to children attending primary education or students just in grade 1 or 2 of primary, the sample sizes are in all except three countries 2000 children.¹⁷

¹⁶ EGRA proposes to rely stronger on hypothesis testing, which, they argue is less demanding in terms of sample size.

¹⁷ When comparing the sample sizes with TIMSS or other school based surveys, it must be considered, that the school based surveys have strong cluster effects and thus smaller effective sample sizes. MICS households surveys are not that strong effected by clustering of the sample, the effective sample size is near to the numbers shown in the table



Now, expecting data of MICS6, which are not yet available, the question is, how many primary school age children can be expected in the sample of questionnaires for children 5 to 15. Since for each household one child is selected, the sample size will decrease, especially since many of the children 5 to 17 are not in primary school, meaning sometimes older children will be sampled, not contributing to the subsample of interest for indicators on primary education. Table 2 provides an overview of sample sizes for a simulated sample of children, as it would have been established with MICS5 samples assuming the procedures planned for MICS6. For example, in Bangladesh, the number of children in primary education sampled for the child questionnaires (5 to 17) would be 15,900, of which 7,400 would be in grade 1 or 2. Also for countries with comparable small samples, such as Kosovo, Kyrgyzstan or Sao Tome and Principe the sample would still be close to or above 1,000 children in primary education. Sufficient to make reliable estimates on the future indicator on language of instruction. Even separation by early and late primary grades will be feasible with the sample.

MICS6 collects data on the language of the children. For the simulation it was assumed that the children sampled speak the same language as the household head, because only this information is available for MICS5. The simulation shows that analysis by individual language group for children in primary will most likely be limited by sample sizes. Already for the second most frequent language group most countries will have less than 500 children in the sample. Estimates on those and smaller language groups will be associated with substantial sampling errors.

In summary, it can be expected that future data from MICS6 will provide solid sample to monitor the use of languages by children attending primary school even for single grades or combinations of two grades. To analysis by single language groups the target sample may be too small.

Country	Households sampled		Children in households		
	total	with children 5 to 17	age 5 to 17	of which: attending primary education	of which: attending grade 1 or 2
Bangladesh	55120	36979	71038	31283	14099
Dominican Republic	33097	16389	28719	20213	5636
El Salvador	14160	7520	13426	6192	2087
Guinea Bissau	6820	5431	16299	9816	4857
Kazakhstan	16791	6675	10981	4032	2122
Kosovo	4870	2472	5375	1859	700
Kyrgyzstan	7190	4170	7945	2650	1424
Nepal	12975	8194	17295	9365	3301
Pakistan (Sindh)	19360	13031	40107	11866	5285
Sao Tome and Principe	3930	2243	4900	2866	912

due to the relative small design effect of about 1.5 to 2.0. This is compared to a design effect in SACMEQ between 4 and 13 (but with the notable exception of Zanzibar and Seychelles).

Turkmenistan	6101	3435	6601	5899	1277
Zimbabwe	17047	10409	22481	13597	-

Table 1: Sample size in MICS 5, selected countries.

Country	Children in <u>simulated</u> sample			
	Attending primary education	Attending in grade 1 or 2	Speaking most frequent language	Speaking second most frequent language
Bangladesh	15921	7357	15362	553
Dominican Republic	10826	3297	10222	511
El Salvador	3364	1184	3360	2
Guinea Bissau	2919	1594	1055	552
Kazakhstan	2414	1274	1686	588
Kosovo	840	334	781	28
Kyrgyzstan	1313	722	1057	165
Nepal	4195	1524	2063	270
Pakistan (Sindh)	3646	1681	1632	644
Sao Tome and Principe	1255	452	969	132
Turkmenistan	3013	690	2600	219
Zimbabwe	6206	6206	4723	982

Table 2: Simulated sample size: For how many children would the child questionnaire be administered given the sampling procedure for MICS6

5 Summary and Recommendations

5.1 Definition: Limit to main languages

The data sources reviewed here for indicator on 4.5.2 focus almost exclusively on monolingual situations for home language and language of instruction. Interpretation is thus limited to statements about main languages of instructions and the match to student's main home language. Indicators calculated on such data would be only sensitive to change if in education systems the main home languages replace other languages as main language of instruction. Expanding multi language education without the shift towards another main language would not be reflected in such indicators.



There is advocacy that the first or home language should be the language of instruction until at least grade 6 e.g. (Ouane & Glanz, 2011, p. 29). This target could be monitored with the existing data sources.

Recommendation:

Using data sources limited to capture home languages and language of instruction monolingual has a strong normative component towards specific language policies. This implicit notion of the indicator should be well explained and interpretation of the potential indicators must be limited to this situation. 3.1.2 shows that instruction in mother tongue complements instruction in the main language of test or instruction in most systems participating in TIMSS/PIRLS. It would therefore be wrong to derive from indicators focusing on main languages whether children have or have not access to instruction in their language at all. The indicators show only if their main language of instruction is different from their main home language. This is especially critical in systems with a transition between languages when only data on higher grades can be analysed.

5.2 Language of instruction

Most data sources are not obtaining information on the language of instruction individually from children but generalized from process data of the assessment used. This can be problematic if languages of instruction are different for different subjects (e.g. Singapore) or language use in classrooms is more flexible.

This is different for the MICS survey. Here language of instruction will be obtained for each individual child. This could serve as an example for other household surveys aiming to capture educational processes.

The EMIS systems and student registers reviewed for this document do not collect explicit data on the language(s) of instruction. Language of instruction seems to be a parameter of the school/class or educational programme.

Recommendation:

Collecting data on languages used in schools could be improved easily by adding few items to the school questionnaire. The item shown in Questionnaire example 13 from PIRLS 2011 is an example that could be considered. It is important to be able to distinguish whether such instruction is available or not. Questionnaire example 12 from PIRLS and TIMSS is thus less useful. National data collections should consider to obtain the data by language used.

A better understanding is needed on how national systems can report on the languages of instruction used in their system and the distribution of students by those languages of instruction. EMIS systems should be reviewed to determine whether a collection of language of instruction at individual or class room level is needed and possible with current systems. If instruction is provided in several languages, the main language for reading and mathematics instruction should be identified. It could be considered to follow Questionnaire example 28 with its focus on language of mathematics and reading instruction.

If only one main language is in use for instruction in a system or if the language of instruction is unique within a given region or school type or educational programme, the collection by class room might



not be necessary. However, EMIS systems should aim to allow the assignment of the language(s) of instruction in use to each school based on data from regions or school type.

Countries with national language policies related to education should be invited to review their EMIS systems to ensure that the implementation of national language policies can be monitored.

In order to provide recommendations for household surveys, the first data from MICS6 should be analysed to verify the feasibility of the MICS approach to capture language of instruction from the interview with guardians and children.

5.3 Definition: Home or first language (mother tongues, ...)

A multitude of similar items on children's languages are used throughout data sources. However, reporting based on international data focus on the current use of languages in the child's home or "out of school" context. Two types of items are most common. One group of data sources records which (one) language is most often used (LLECE; EGRA, MICS), the other group of items records how often a given reference language (the language of test or instruction) is used by children (TIMSS/PIRLS, PASEC, SACMEQ).

For the purpose of 4.5.2 both types of items allow to determine whether the main language of instruction (or test) is also the current main language of the child. Therefore, it seems that both items can be used to reflect on a similar construct (match of language of instruction and home language).

Items on mother tongue or "first languages learned/spoken before school" are not as common in the surveys and are also not constantly used over different survey rounds. The reliability of students reporting on the languages seems also limited (3.2.4).

National administrative data use mainly different but similar concepts of home language. The gathering of data on the mother tongue of students seems to be an exception.

The (United Nations, 2008, p.138) recommends not to overstate the need for international comparisons on the language of individual persons. This recommendation should be kept in mind when advising countries on survey studies. A balance should be ensured between national needs and specific situations requiring international comparisons.

Recommendation:

Recommendations for future international survey should further foster the use of one concept to capture children's language to support harmonizing reporting. Items capturing the current language use of children should be promoted, rather than mother tongue or first learned language. Here the most common concept, use of language at home, should be given preference over questions on language use in everyday life or out of school. This is simply to mainstream items across different surveys and to reduce use of unnecessary similar concepts.

No clear recommendation should be given on the question which group of items to use, either "Which language is spoken at home" or "How frequent the language of instruction or test is spoken". Most likely is the decision closely related with the original purpose of the data collection. Since both concepts lead to a similar construct with respect to reporting of 4.5.2, a recommendation seems not necessary.

For national EMIS systems and student registers, home language seems to be the most widely used definition. Countries expanding their national statistical system should be invited to use a concept of



home language except where specific national circumstances support other definitions. Ideally, the main home language or languages should be recorded and reported. For EMIS systems reporting data aggregated by classes or schools by individual languages is likely to be ambitious. Here a short list of languages to be considered could be used. It is important that it is possible to reconstruct the relationship between students' languages to the language used in class.

When defining 4.5.2 to be based on home language, alternative data based on first language or mother tongue should still be accepted when without alternative. But this should be marked clearly.¹⁸ It should be avoided to report trends switching between concepts.

5.4 Frequency of language use

Items recording the frequencies of home language use differ with respect to the categories used. Earlier rounds of PIRLS used three categories, TIMSS used in recent rounds four categories. It seems that using four categories allows better differentiation of a main language for bi-/multilingual homes (3.2.1).¹⁹

For the purpose of reporting on 4.5.2 it could be debated, whether a child should be classified as speaking a given language of instruction at home if it speaks it sometimes (e.g. half of the time with a second language) or only if the language is used always or almost all of the time. SACMEQ seems to classify "sometimes" as speaking the language, while GEM and TIMSS/PIRLS reporting set a cut between to report "sometimes" as not speaking the language.

If the goal for reporting is to pool different data sources using different items, the decision on where to set the cut should be made in a way that allows to reach similar constructs when including indicators derived from MICS/LLECE or EGRA and from PIRLS/TIMSS, SACMEQ and PASEC. Therefore, the cut should be set in a way that deriving the use of language of instruction from "what is the language you speak most at home" comes conceptually as close as possible to answers derived from "How often do speak the language of instruction". It seems that this goal would be better achieved when cut between "almost always" and "sometimes". However, the price is that more bilingual children would be reported as having no access to education in a language they speak.

It should be noted that the interpretation of data is up to each analyst. The purpose is not to recommend using the same split in all analytical work. But recommendations should support ways of phrasing the items, that allow other researcher, when pooling data, to set most consistent cuts.

Recommendation:

Items on the frequency of language use should have at least four categories or categories should be phrased in a way, that for bilingual students, speaking a given language most frequent, it is clear whether the language of instruction is the main language at home. For international pooling of data children should be classified as speaking the language of instruction at home, when they speak it

¹⁸ As it is common to report on education expenditure with a mark when using public expenditure only to instead of public and private.

¹⁹ The assessment studies were mainly designed to best identify factors associated with the tested domain. Here the focus is less on descriptive statistics than on explaining variance in achievement. Therefore, the observation made here is not on the purposefulness of the items in the context of the study, but on there properties when deriving language statistics.



always or almost always. This is to reach best consistency with the alternative items asking for a main language and not frequency of use.

5.5 Inconsistencies in reporting by parents and children

As for any background information from surveys for children of young ages, the question arises whether children can provide reliable answers to questions on their families and educational context. That is especially problematic for questions on the income or occupation of parents. Some surveys obtain therefore information from parents to complement data obtained from students. Some of the information is obtained twice, allowing to compare students and guardians report. Figure 4 and 5 show substantial inconsistencies between reporting by students and parents. This raises the question, whether recommendations should be given towards which data to prefer when both sources (parents/ guardians and students) are available or when new surveys are to be designed. When discussing this question, it must be kept in mind that statistical information, especially when obtained from young children, will always be subject to measurement error.

Recommendation:

Additional data collections from parents within school based surveys are always increasing costs and complexity of surveys. For the purpose of international monitoring, no recommendation should be given that increases costs of surveys if the survey responsible agencies deem themselves the addition unnecessary.

If both sources of data are available, an analysis of missing data for both sources should be done. Typically, data from parents have higher missing rates due to non-participation of parents. The administration is simply more complex and not under control of the test administrators. Missing rates are mostly biased with respect to socio-economic factors. The source with substantially lower missing rates should be used, if a difference is observed.

Since measurement error is quite likely not unbiased, trend data should draw for different cycle of the same survey on the same source (parents or students) and not switch sources over years and present trends.

If it is the most common situation that data are based on student responses, it should be indicated if the source of information is different, i.e. guardians.

5.6 Total primary education as target level

The indicators on 4.5.2 should cover children in primary education. Data show that the use of home languages in instruction is more frequent in lower grades than in higher grades (3.1.2). This is especially the case where language policies exist promoting a transition of language of instruction from early to higher grades.

Therefore, data from studies on higher grades alone cannot provide a complete picture. The percentage of students in total primary education whose first or home language in a reference year is the language of instruction would be underestimated. Similarly, the use of data on lower grades would overestimate the share. Cross country comparisons would be biased by the data source used.



In addition, a focus on higher (or lower) grades only would be biased the picture because students speaking other languages than the languages of instruction are more likely to drop out.

The use of EMIS and register data on languages would allow reporting on all students in the system and overcome the problems related to data sources with data available only for specific grades. Yet, this requires for systems with different languages of instruction a good recording of each language of instruction

Recommendation:

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

Reporting of 4.5.2 should be done separately for early grades (ideally grade 1 to 3) and grades towards the end of primary education. This would also be in line with the proposed reporting on SDG indicator 4.1.1 and 4.1.2.

If not sufficient data are available, reporting could be limited to higher primary grades, but in that case indicators should be clearly labelled as showing “the percentage of students having education provided mainly in their home language up to the end of primary”.

Summative indicators on total primary must be derived from data on lower and higher grades separately for countries with multi language education. For regional or global reporting, the calculation of regional figures must make provision for cases where data are available only for lower or higher grades. Simply generalizing from one grade of primary to all grades is not appropriate. For countries having identical policies on language of instruction for all grades, this approach might be acceptable.

Surveys administering school questionnaires could include simple items on language use in instructions in grade 1 to 3. This could help in the estimation of indicators for the whole grade.

5.7 Data quality and precision

The discussed international data sources have sufficient sample sizes and sound sampling processes to produce reliable estimates on their target populations. Sampling errors are not an important limitation when using the data. Only for the EGRA project seems sample size and sample process problematic for some implementations in some countries.

However, a more critical issue is that the target populations are in many cases biased when it comes to language issues. Students not speaking the language and students in schools with different languages of instruction tend to be excluded from many large scale assessment surveys for schools. This is different for the households survey MICS, where sampling is independent from school settings and thus not limited by issues of languages used in test situations. Therefore, if the forthcoming MICS6 lives up to the expectations it seems the most valid source covering all children. However, also for MICS the quality and completeness of the sample needs to be questioned for each implementation.

Recommendation:

Data from MICS 6 should be analysed as soon as available to understand their quality. For reporting on primary children’s language MICS seems to provide the least biased sample.



For school based sample surveys a good documentation of the sample exclusions is needed to judge the quality and appropriateness of the data. It might be necessary to adjust the results from those studies to account for excluded populations. This depends on a judgement of the size of exclusion. Adjustments are especially needed, when it is the nature of the exclusion that the children do not speak the language of instruction and can thus not be tested or when exclusions are simply because for schools of special language groups is no test material available.

5.8 Detailed information on different languages

About half of the data sources reviewed allow to understand the distribution of children by their main home languages. Yet, some important data sources do not allow this.

Recommendation:

International reporting on 4.5.2 should be focused on the core phenomena, the match/mismatch of language of instruction and home language. In order of data parsimoniousness, collection of data by language groups should not be pushed for as long as the maintainer of the original data sources do not see the need for it. However, for policy making at the national level disaggregating data by national relevant language groups seems necessary and should be advised.

5.9 Comments on principles for Global Monitoring Indicators

Leadership Council of the Sustainable Development Solutions Network (2015, p. 17) identified ten principles for SDG indicators. They are formulated for the global SDG indicators and thus not as strict to be applied for thematic indicators, but most provide a good guideline for summarizing the discussion on 4.5.2.

1. Limited in number and globally harmonized

Many different items are used by the reviewed data sources (See Annex). Yet there is an intersection amongst most background questionnaires in use: All allow to observe whether the main language of test matches the main home language. Focusing on this one indicator would make global monitoring possible. Other concepts should not be used for regional and international reporting to limit the number of indicators and avoid parallelism of concepts. National and some regional work will and should go into more detail concerning languages, but globally this should be avoided because of the principle.

2. Simple, single - variable indicators, with straightforward policy implications

The current data sources allow mainly to report with a monolingual perspective on home languages and language of instruction. The messages drawn from the indicators must consider this perspective. Giving such importance to just the main language of instruction has maybe implicit policy implications. This should be discussed and be subject to an conscious decision.



3. Allow for high frequency monitoring

With the importance of SDG indicator 4.1.1, the availability of international and national assessment data will grow. Most international projects produce data in cycles of 4 to 5 years. Though not annual, the data will allow for the observation of change. Considering the slow pace of change and the standard errors of the assessments, more regular assessments seem not necessary.

High frequency monitoring has its purpose only if the indicators are sufficient sensitive for change. With the limit to report only on main languages of instruction, the indicator is not sensitive to improved access to language in the home language when only complementing instruction in another language of instruction.

4. Consensus based, in line with international standards and system - based information

As mentioned above, the notion of a single main language of instruction and main home languages has implications on interpretation of the data. It should be clarified whether the interpretation of such indicators are in line with international standards. The fact that most international assessment studies choose that route does not mean it is built on a consensus of expert in language policies. The consensus amongst assessment experts is likely due to the fact that the chosen items are very powerful in explaining achievement.²⁰

5. Constructed from well - established data sources

The discussed data sources allow to cover nearly 120 countries in the world²¹. The considered data sources (MICS, PIRLS/TIMSS, SACMEQ; PASEC and LLECE) are well established and sound survey projects with a big international community of researchers working with the data and ensuring the quality. This principle can certainly be matched.

6. Disaggregated

All discussed data sources are available as micro-data. All allow for relevant disaggregation of the data, such as characteristics of the individual child (e.g. sex, attainment of parents, migration²²), economic activity and spatial dimensions (urban and rural).

²⁰ This point becomes clear when considering that most survey ask children for the amount of books in their home. That is simply because this is a powerful proxy for cultural capital, which is know to explain a substantial fraction of variance.

²¹ This number assumes that most countries joining MICS6 will administer the questionnaire for children aged 5 to 17.

²² According to Hooper et al. (2015) PIRLS 2016 does not ask for the migration background of the family anymore and deviates from previous TIMSS questionnaires. This would be a dramatic gap to better understand the context of children not speaking the language of instruction.

7. Universal

The issue of language of children's access to education in their home language is certainly.

8. Mainly outcome - focused

With respect to the goals of education, the indicator is rather input than outcome focused. Yet, as in other cases, input metrics can play a critical role in driving and tracking the changes needed for sustainable development. This seems to apply for this indicator, assuming consensus amongst experts that language policies are one key to better learning.

ANNEX A: Example questionnaire Items

5.10 PIRLS / TIMSS

Which languages did you learn to speak when you were little?

If you learned more than one language at the same time when you were little you can check "Yes" for more than one language.

	Yes	No
	↓	↓
a) <language of test>	<input type="radio"/>	<input type="radio"/>
b) <country-specific>	<input type="radio"/>	<input type="radio"/>
c) <country-specific>	<input type="radio"/>	<input type="radio"/>
d) <country-specific>	<input type="radio"/>	<input type="radio"/>
e) <country-specific>	<input type="radio"/>	<input type="radio"/>
f) Other	<input type="radio"/>	<input type="radio"/>

Please write in the name of the language.

Questionnaire example 1: Student questionnaire, PIRLS 2001: First language(s).

Which language did you speak before you started school?

If you spoke more than one language at the same time before you started school you can check "Yes" for more than one language.

- | | Yes | No |
|-----------------------------|-----------------------|-----------------------|
| a) <language of test> | <input type="radio"/> | <input type="radio"/> |
| b) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| c) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| d) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| e) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| f) Other | <input type="radio"/> | <input type="radio"/> |

Questionnaire example 2: Student questionnaire, PIRLS 2006: First language(s).

What language did your child speak before he/she began school?

If your child spoke more than one language check "Yes" for more than one language.

Check **one** circle for each line.

- | | Yes | No |
|-----------------------------|-----------------------|-----------------------|
| a) <language of test> | <input type="radio"/> | <input type="radio"/> |
| b) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| c) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| d) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| e) <country-specific> | <input type="radio"/> | <input type="radio"/> |
| f) Other | <input type="radio"/> | <input type="radio"/> |

Note: Almost identical Question used in PIRLS/TIMSS 2011, PIRLS 2006.

Questionnaire example 3: Home (parent) questionnaire, TIMSS 2015: First language(s) of child.

**How often do you speak <language of test> at home?**

Fill one circle only.

I always speak
<language of test> at home --

I almost always speak
<language of test> at home --

I sometimes speak <language
of test> and sometimes speak
another language at home --

I never speak <language
of test> at home --

Note: Almost identical Question used in TIMSS 2007.

Questionnaire example 4: Student questionnaire, TIMSS 2015: Home language.

How often do you speak <language of test> at home?

Fill one circle only.

Always or Almost Always --

Sometimes --

Never --

Questionnaire example 5: Student questionnaire, PIRLS 2001: Home language.

How often do you speak <language of test> at home?

Fill one circle only.

- I always speak <language of test> --
- I sometimes speak <language of test>
and sometimes speak another language -
- I never speak <language of test> --

Questionnaire example 6: Student questionnaire, PIRLS 2006: Home language.

How often do you speak <language of test> at home?

Fill one circle only.

- I always or almost always
speak <language of test> at home --
- I sometimes speak <language
of test> and sometimes speak
another language at home --
- I never speak <language
of test> at home --

Questionnaire example 7: Student questionnaire, PIRLS/TIMSS 2011: Home language.

How often do you speak <language of test> with adults living in your home?

Fill one circle only.

- Always or Almost Always --
- Sometimes --
- Never --

Questionnaire example 8: Student questionnaire, PIRLS 2001: Home language spoken with adults.

When talking at home with your child, what language does the child's father (or stepfather or male guardian) use most often? What language does the child's mother (or stepmother or female guardian) use most often?

Check **one** circle in each column.

	Child's father	Child's mother
a) <language of test> -----	<input type="radio"/>	<input type="radio"/>
b) <country-specific> -----	<input type="radio"/>	<input type="radio"/>
c) <country-specific> -----	<input type="radio"/>	<input type="radio"/>
d) <country-specific> -----	<input type="radio"/>	<input type="radio"/>
e) <country-specific> -----	<input type="radio"/>	<input type="radio"/>
f) Other -----	<input type="radio"/>	<input type="radio"/>
g) Not applicable -----	<input type="radio"/>	<input type="radio"/>

Note: Identical Question used in PIRLS/TIMSS 2011.

Questionnaire example 9: Home (parent) questionnaire, PIRLS 2006: Home language(s).

When talking at home with your child, what language does the child's father (or stepfather or male guardian) use? What language does the child's mother (or stepmother or female guardian) use?

Check all that apply.

	Child's father	Child's mother
a) <language of test> -----	<input type="checkbox"/>	<input type="checkbox"/>
b) <country-specific> -----	<input type="checkbox"/>	<input type="checkbox"/>
c) <country-specific> -----	<input type="checkbox"/>	<input type="checkbox"/>
d) <country-specific> -----	<input type="checkbox"/>	<input type="checkbox"/>
e) <country-specific> -----	<input type="checkbox"/>	<input type="checkbox"/>
f) Other -----	<input type="checkbox"/>	<input type="checkbox"/>
g) Not applicable -----	<input type="checkbox"/>	<input type="checkbox"/>

Questionnaire example 10: Home (parent) questionnaire, TIMSS 2015: Home language(s).

How often does your child speak <language of test> at home?

Check one circle only.

Always ---

Almost always ---

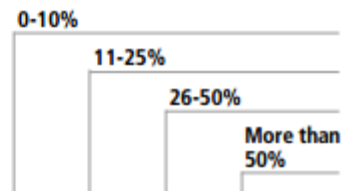
Sometimes ---

Never ---

Questionnaire example 11: Home (parent) questionnaire, TIMSS 2015: Use of language of test at home.

Approximately what percentage of students in your school ...

Check **one** circle for each line.



d) receive some instruction at school in their home language (other than <language of test>)

Questionnaire example 12: School questionnaire, PIRLS 2001: Instruction in languages other than test-language.

For students in <fourth grade> and below, does your school make provisions for reading instruction in mother tongue for students whose mother tongue is not <language of test>?

Check **one** circle only.

Yes---

No---

Questionnaire example 13: School questionnaire, PIRLS 2011: Instruction in languages other than test-language.



5.11 SACMEQ

4. Do you speak English outside school?
(Please tick *only one* box.)

PENGLISH

- (a) Never
 (b) Sometimes
 (c) Most of the time
 (d) All of the time

Source: (SACMEQ, 2007, p. 3)

Questionnaire example 14: Student questionnaire, SACMEQ 3: First language(s).

4. Do you speak English outside school?
(Please tick *only one* box.)

- (1) Never
 (2) Sometimes
 (3) All of the time

Source: (SACMEQ, 2000, Botswana)

Questionnaire example 15: Student questionnaire, SACMEQ 2: First language(s).

5.12 PASEC

VOTRE ENSEIGNEMENT

27. Utilisez-vous la langue maternelle parlée par la majorité des enfants pour vous faire comprendre pendant les cours ?

(Veuillez ne cocher qu'une seule case)

- Toujours ou presque Souvent Occasionnellement,
parfois Jamais ou presque
jamais

Note: the item is used identically for grade 2.

Questionnaire example 16: Teacher Questionnaire, 6th grade, PASEC 2014, Language used in instruction (Example: Burundi)

15. Quelle langue parles-tu chez toi ?

Je parle toujours le français

Je parle parfois le français et une autre langue

Je ne parle jamais le français

Note: the item is used identically for grade 6.

Questionnaire example 17: Student Questionnaire, 6th grade, PASEC 2014, Frequency of use of language of instruction at home

5.13 LLECE: SERCE

6 ¿Cuál es su lengua materna?

→ Marque con una **X** sólo una.

1. Castellano (portugués)

2. Lengua extranjera

3. Lengua indígena 1

4. Lengua indígena 2

5. Otra lengua indígena. Especifique:

7 ¿Cuál es la lengua que usa mayormente en clase?

→ Marque con una **X** sólo una.

1. Castellano (portugués)

2. Lengua extranjera

3. Lengua indígena 1

4. Lengua indígena 2

5. Otra lengua indígena. Especifique:

**Questionnaire example 18: Teacher questionnaire, SERCE: Mother tongue of teacher himself.
Language mainly spoken in class.**

DRAFT

3 ¿En qué idioma habla tu profesor en clase la mayor parte del tiempo?

→ Marca con una **X** sólo una.

1. Castellano o portugués
2. Lengua extranjera
3. Lengua indígena 1
4. Lengua indígena 2
5. Otra lengua indígena

4 ¿Cuál es el idioma en que hablas en tu casa?

→ Marca con una **X** sólo una.

1. Castellano o portugués
2. Lengua extranjera
3. Lengua indígena 1
4. Lengua indígena 2
5. Otra lengua indígena

Questionnaire example 19: Student questionnaire grade 6, SERCE: Language used by teacher in instruction, language used at home

5 ¿Qué idioma aprendió a hablar primero el estudiante?

→ Marque con una **X** sólo una.

1. Castellano o portugués
2. Lengua extranjera
3. Lengua indígena 1
4. Lengua indígena 2
5. Otra lengua indígena.

Questionnaire example 20: Home questionnaire, SERCE: First language spoken by student



5.14 EGRA

5.		<p>[Language 1]..... 1</p> <p>[Language 2]..... 2</p> <p>[Language 3]..... 3</p> <p>Other (specify):..... 4</p> <p>_____</p>
	What is your native language?	

Questionnaire example 21: Teacher questionnaire, EGRA (within Snapshot of School Management Effectiveness): Native language of teacher

8.		<p>[language]..... 1</p> <p>English..... 2</p> <p>Other (specify):..... 3</p> <p>_____</p>
	What language do you speak most often at home?	

Questionnaire example 22: Student questionnaire, EGRA (within Snapshot of School Management Effectiveness): Home language of student

**Questionnaire example 24: PISA 2015, student questionnaire (age 15): language spoken at home most often.**

5.16 MICS

<p>FL7. Which language do you speak most of the time at home?</p> <p><i>Probe if necessary and read the listed languages.</i></p>	<p>LANGUAGE 11</p> <p>LANGUAGE 22</p> <p>LANGUAGE 33</p> <p>OTHER (<i>specify</i>) 6</p> <p>DK8</p>
<p>FL9. What language do your teachers use most of the time when teaching you in class?</p> <p><i>Probe if necessary and name the listed languages.</i></p>	<p>LANGUAGE 11</p> <p>LANGUAGE 22</p> <p>LANGUAGE 33</p> <p>OTHER (<i>specify</i>) 6</p> <p>DK8</p>

Questionnaire example 25: MICS6 Questionnaire for children age 5-17: Home language of children. Language used in teaching of child.

DRAFT for

5.17 National Assessment Austria

11. In welcher Sprache hat Ihr Kind zuerst sprechen gelernt?

Wenn Ihr Kind gleichzeitig zwei oder mehr Sprachen sprechen gelernt hat, bitte alle zutreffenden Kästchen ankreuzen.

Deutsch	<input type="checkbox"/>	Rumänisch	<input type="checkbox"/>
Albanisch	<input type="checkbox"/>	Russisch	<input type="checkbox"/>
Arabisch	<input type="checkbox"/>	Serbisch	<input type="checkbox"/>
Bosnisch	<input type="checkbox"/>	Slowakisch	<input type="checkbox"/>
Englisch	<input type="checkbox"/>	Slowenisch	<input type="checkbox"/>
Kroatisch	<input type="checkbox"/>	Tschechisch	<input type="checkbox"/>
Kurdisch	<input type="checkbox"/>	Türkisch	<input type="checkbox"/>
Polnisch	<input type="checkbox"/>	Ungarisch	<input type="checkbox"/>
Romanes	<input type="checkbox"/>	eine andere Sprache	<input type="checkbox"/>

11. Was ist deine Muttersprache?

Muttersprache ist die Sprache, in der du als Kind zuerst sprechen gelernt hast.

Wenn du als Kind gleichzeitig zwei oder mehr Sprachen gelernt hast, kannst du alle zutreffenden Kästchen ankreuzen.

Deutsch	<input type="checkbox"/>
Albanisch	<input type="checkbox"/>
Bosnisch	<input type="checkbox"/>

Questionnaire example 26: Austria Home & Student Questionnaire: First language(s) of child reported by parents and child

DRAFT

12. Wie oft sprechen Sie als Eltern/Erziehungsberechtigte zu Hause mit Ihrem Kind Deutsch?

Bitte **das Zutreffende** ankreuzen.

- Wir sprechen zu Hause immer oder fast immer Deutsch.
- Wir sprechen zu Hause manchmal Deutsch und manchmal eine andere Sprache.
- Wir sprechen zu Hause nie Deutsch.

12. Wie oft sprichst du zu Hause Deutsch?

Bitte **nur ein Kästchen** ankreuzen.

- Ich spreche zu Hause immer oder fast immer Deutsch.
- Ich spreche zu Hause manchmal Deutsch und manchmal eine andere Sprache.
- Ich spreche zu Hause nie Deutsch.

Questionnaire example 27: Austria Home & Student Questionnaire: Home language (Use of language of instruction at home) reported by parents and child

5.18 Living Standards Measurement Study (LSMS)

11. What language is used for...

**PUT LANGUAGE
CODES HERE**

...Mathematics instruction?

...Reading instruction?

...General conversation?

Source: (Grosh & Glewwe, 2000, Vol. 3, p. 52)

Questionnaire example 28: LSMS school questionnaire template: language of instruction



ANNEX B: Statistics and selected results.

Annex B - Table 1: Share of students with other home language than language of test who are in schools that offer mother tongue reading instruction

Country	%Students with other home language in schools offering reading instruction in mother tongue for them	SE
Armenia*	55,92	4,48
Australia	36,08	3,43
Austria	29,35	4,09
Azerbaijan	72,47	4,02
Bahrain*	71,48	4,17
Belgium (Flemish)*	24,58	3,52
Belgium (French)	34,75	5,56
Botswana	50,91	5,32
Botswana+	57,43	5,73
Bulgaria	6,51	1,77
Canada	39,51	3,16
Chile*	28,21	3,75
Colombia	11,96	2,89
Colombia+	11,99	2,90
Croatia	50,93	4,44
Czech Republic	10,52	2,50
Denmark	16,49	2,75
England	21,96	4,55
Finland	33,52	4,10
France	34,22	4,37
Georgia	73,64	3,63
Germany	22,49	3,37
Honduras	27,20	4,63
Hong Kong	24,84	3,87
Hungary	11,60	2,86
Indonesia	40,91	5,19
Iran (Islamic Republic)	11,89	2,58
Ireland	19,21	2,99
Israel	40,53	4,70



Country	%Students with other home language in schools offering reading instruction in mother tongue for them	SE
Italy	26,77	3,43
Japan*	26,86	3,25
Kazakhstan*	62,27	4,27
Korea (Republic)*	6,26	2,33
Kuwait	21,29	4,89
Lithuania	53,35	4,11
Malta	52,58	2,03
Morocco	41,13	6,09
Netherlands	3,43	1,77
New Zealand	24,10	3,39
Northern Ireland	23,32	4,30
Norway	23,90	4,35
Oman	52,10	3,83
Poland	3,11	1,56
Portugal	43,71	4,90
Qatar	68,18	4,01
Romania	20,52	3,81
Russian Federation	26,08	3,66
Saudi Arabia	56,94	4,44
Serbia*	20,55	3,62
Singapore	100,00	0,00
Slovakia	15,33	2,96
South Africa	44,84	6,20
South Africa+	45,66	4,26
Spain	47,43	3,89
Sweden	69,74	4,56
Taiwan (China)	20,68	3,23
Thailand*	17,19	3,47
Trinidad/Tobago	19,76	3,44
Tunisia*	28,71	4,28
Turkey*	3,10	1,12
United Arab Emirates	61,79	2,56
USA	43,86	2,62



SE: Standard error. Source: PIRLS 2011, * TIMSS 2011, +prePirls 2011. Authors own calculations.
Calculated using Jackknife methods with the R-package BIFESurvey.

DRAFT for discussion

Annex B - Table 2: Students report on the use of test language at home

Country	Students speak test language at home:					
	(almost) always	SE	sometimes	SE	never	SE
Algeria*	56,0	2,4	31,8	1,9	12,2	1,0
Armenia*	94,8	0,6	3,7	0,4	1,4	0,4
Australia	84,9	1,1	13,8	1,0	1,4	0,2
Austria*	87,6	0,7	10,0	0,6	2,4	0,3
Bahrain	67,9	0,8	27,5	0,6	4,5	0,5
Belgium (Flemish)	78,6	1,0	18,4	0,9	3,0	0,5
Bulgaria	76,1	2,3	17,7	1,7	6,2	1,0
Canada	75,3	0,9	21,9	0,8	2,9	0,3
Chile	89,7	0,6	5,8	0,4	4,5	0,4
Colombia*	88,8	0,9	8,1	0,8	3,0	0,3
Croatia	92,2	0,7	7,0	0,5	0,8	0,4
Cyprus	75,2	1,2	21,4	0,9	3,4	0,4
Czech Republic	91,5	0,4	7,8	0,4	0,6	0,1
Denmark	88,4	0,8	10,9	0,7	0,7	0,1
El Salvador*	92,6	0,8	5,1	0,6	2,3	0,3
England	82,8	1,5	15,6	1,4	1,7	0,3
Finland	88,9	0,7	9,8	0,7	1,3	0,2
France	83,4	0,8	15,7	0,8	0,8	0,1
Georgia	87,5	0,9	11,7	0,8	0,8	0,5
Germany	80,6	1,0	18,3	1,0	1,1	0,2
Hong Kong	70,5	1,7	28,2	1,7	1,2	0,3
Hungary	97,6	0,3	2,2	0,3	0,2	0,1
Indonesia	43,8	1,9	42,2	1,7	14,1	1,0
Iran (Islamic Republic)	66,7	1,9	16,6	1,1	16,7	1,5
Ireland	87,5	0,8	10,3	0,7	2,2	0,3
Italy	84,0	0,8	13,8	0,8	2,3	0,3
Japan	98,4	0,2	1,4	0,2	0,2	0,1
Kazakhstan	87,4	1,1	11,6	1,0	1,1	0,2
Korea (Republic)	91,7	0,5	8,2	0,5	0,1	0,0
Kuwait	27,8	1,6	34,7	1,5	37,4	1,7
Latvia*	88,5	1,5	8,9	1,1	2,6	0,6

Country	Students speak test language at home:					
	(almost) always	SE	sometimes	SE	never	SE
Lithuania	90,8	0,5	8,9	0,5	0,3	0,1
Mongolia*	86,4	1,3	11,8	1,1	1,7	0,3
Morocco	46,7	1,8	24,5	1,4	28,8	1,8
Netherlands	80,1	1,6	17,1	1,2	2,8	0,7
New Zealand	84,0	0,9	14,3	0,8	1,7	0,2
Northern Ireland	91,8	1,0	6,9	0,9	1,3	0,2
Norway	84,9	1,2	13,6	1,1	1,5	0,2
Oman	62,9	1,4	24,5	1,0	12,7	1,1
Poland	96,1	0,4	3,6	0,3	0,3	0,2
Portugal	91,2	0,5	8,1	0,5	0,7	0,1
Qatar	53,0	1,3	39,1	1,1	7,8	0,7
Russian Federation	90,6	1,1	8,4	1,0	1,0	0,3
Saudi Arabia	79,1	1,4	13,3	1,0	7,6	0,8
Scotland*	91,2	0,8	6,0	0,5	2,7	0,6
Serbia	94,6	0,8	4,9	0,7	0,5	0,2
Singapore	48,6	0,6	47,7	0,6	3,7	0,3
Slovakia	84,6	1,3	12,0	1,0	3,4	0,7
Slovenia	86,1	1,0	11,0	0,8	2,9	0,4
Spain	73,3	1,3	17,7	0,7	9,1	0,9
Sweden	83,6	1,3	14,9	1,1	1,5	0,3
Taiwan (China)	59,3	1,0	39,7	0,9	1,0	0,2
Tunisia*	26,1	1,7	48,8	2,0	25,2	1,8
Turkey	84,7	1,2	14,0	1,0	1,3	0,3
Ukraine*	74,0	2,1	20,7	1,7	5,3	0,6
United Arab Emirates	53,5	0,9	40,0	0,8	6,4	0,3
USA	78,9	1,0	19,1	0,9	2,1	0,2
Yemen*	84,5	1,7	11,2	1,3	4,2	0,9

SE: Standard error. Source: TIMSS 2015, * TIMSS 2007. Authors own calculations. Calculated using Jackknife methods with the R-package BIFIEsurvey.

Annex B - Table 3: Comparison of two ways to ask for home language: Students who report to speak language of instruction always (or almost always) (%)

	TIMSS/PIRLS 2011	SE	TIMSS 2007	SE	PIRLS 2006	SE
Armenia	81,2	0,9	94,8	0,6	m	
Australia	79,5	1,3	90,5	1,0	m	
Austria	76,6	1,1	87,6	0,7	73,8	1,2
Azerbaijan	88,4	1,3	m		m	
Bahrain	53,7	1,3	m		m	
Belgium (Flemish)	75,0	1,3	m		76,8	1,3
Belgium (French)	71,4	1,4	m		66,1	1,3
Botswana	10,4	0,8	m		m	
Bulgaria	75,1	2,1	m		69,7	2,2
Camda	73,7	0,9	m		m	
Chile	80,0	0,7	m		m	
Colombia	88,5	0,9	88,8	0,9	m	
Croatia	85,5	1,0	m		m	
Czech Republic	86,8	0,6	97,1	0,3	m	
Denmark	81,9	0,9	93,6	0,9	81,5	1,1
England	79,2	1,7	93,0	0,6	75,5	1,3
Finland	89,1	0,7	m		m	
France	77,6	1,3	m		65,5	1,2
Georgia	81,7	1,2	91,8	0,7	85,2	1,5
Germany	80,4	1,1	92,1	0,6	72,8	1,0
Honduras	91,7	1,0	m		m	
Hong Kong	68,2	0,9	81,9	0,9	65,5	0,8
Hungary	97,4	0,3	97,8	0,4	75,2	1,2
Indonesia	34,9	2,5	m		38,3	2,0
Iran (Islamic Republic)	54,4	1,5	62,3	2,1	53,0	2,2
Ireland	84,4	0,8	m		m	
Israel	74,9	1,1	m		57,5	1,1
Italy	79,4	1,0	96,4	0,2	70,8	1,2
Japan	85,2	0,7	98,7	0,2	m	
Kazakhstan	78,3	1,4	93,0	1,3	m	
Korea (Republic)	75,2	0,9	m		m	



	TIMSS/PIRLS 2011	SE	TIMSS 2007	SE	PIRLS 2006	SE
Kuwait	48,6	1,8	73,8	1,8	25,6	1,5
Lithuania	82,8	1,2	97,5	0,4	79,2	1,0
Malta	17,1	0,6	m		m	
Morocco	24,3	1,8	49,7	2,6	49,5	2,9
Netherlands	77,9	0,8	89,4	1,2	75,7	1,2
New Zealand	74,3	1,3	87,2	0,8	73,3	1,0
Northern Ireland	91,2	0,6	m		m	
Norway	81,5	1,1	94,3	0,5	79,8	0,9
Oman	59,1	1,7	m		m	
Poland	90,8	0,5	m		85,4	0,6
Portugal	89,7	0,7	m		m	
Qatar	46,2	1,8	71,4	0,6	61,2	0,7
Romania	89,4	1,5	m		81,3	1,6
Russian Federation	84,6	1,6	91,6	1,4	81,9	1,1
Saudi Arabia	69,8	2,1	m		m	
Serbia	90,0	0,7	m		m	
Singapore	32,4	0,5	50,3	0,9	21,2	0,7
Slovakia	79,6	1,6	86,7	1,5	71,3	1,6
Slovenia	m		89,6	0,8	m	
South Africa	42,6	2,2	m		62,1	1,2
Spain	68,7	1,1	m		60,5	1,3
Sweden	80,1	1,0	91,6	1,0	75,1	1,2
Taiwan (Chim)	50,1	0,9	84,3	0,8	36,2	0,8
Thailand	53,7	2,5	m		m	
Trinidad/Tobago	83,0	1,2	m		77,5	1,3
Tunisia	18,3	1,7	26,1	1,7	m	
Turkey	75,5	1,4	m		m	
United Arab Emirates	46,9	0,9	m		m	
USA	85,9	0,6	86,8	0,8	71,5	1,4
Yemen	80,9	1,8	84,5	1,7	m	

SE: Standard error. Source: TIMSS, PIRLS. Authors own calculations. Calculated using Jackknife methods with the R-package BIFIEsurvey.

Annex B - Table 4: Students and guardians report on the first language of the student: Dis-/agreement between respondents

Note: 100% represent the children for who at least one questionnaire (home or student) indicates that they did not learn the language. Children with consistent reporting that they learned the language of instruction are not considered.

Country	Student reports first language different from language of instruction / guardians disagree (%)	SE	Agreement: Student and Guardian report first language different from language of instruction (%)	SE	Guardian reports first language different from language of instruction / student disagrees (%)	SE
Austria	31,0	3,5	37,9	4,6	31,1	3,3
Belgium (Flemish)	33,2	4,1	37,2	6,4	29,6	3,8
Belgium (French)	44,1	3,8	27,6	4,5	28,3	3,9
Bulgaria	23,1	4,2	41,0	7,6	35,9	4,9
Denmark	47,1	7,1	25,6	5,2	27,3	4,8
England	37,2	5,8	36,2	9,1	26,5	8,3
France	54,6	6,3	20,8	4,0	24,6	3,2
Georgia	38,8	8,6	41,1	24,3	20,2	4,0
Germany	36,7	4,4	24,1	4,0	39,2	5,0
Hong Kong	56,8	5,6	15,0	3,7	28,2	4,1
Hungary	78,4	14,7	7,4	4,7	14,2	5,5
Iceland	75,7	10,8	13,1	4,6	11,2	3,7
Indonesia	56,7	6,8	6,5	1,9	36,7	6,3
Iran (Islamic Republic)	14,0	2,4	71,0	5,2	15,0	2,4
Israel	31,2	4,1	20,6	3,5	48,2	7,2
Italy	0,0	0,0	68,7	8,7	31,3	7,2
Kuwait	12,5	0,9	44,5	2,3	43,0	2,1
Latvia	30,5	7,9	17,9	6,4	51,6	6,9
Lithuania	64,7	12,5	15,6	5,7	19,7	4,9
Luxembourg	15,0	0,5	64,0	0,7	21,0	0,6
Macedonia (FYR)	37,2	7,0	28,6	8,6	34,2	7,4
Moldova (Republic)	45,1	9,4	12,1	2,8	42,8	7,3
Morocco	35,9	6,3	30,2	11,7	33,9	7,5



Country	Student reports first language different from language of instruction / guardians disagree (%)	SE	Agreement: Student and Guardian report first language different from language of instruction (%)	SE	Guardian reports first language different from language of instruction / student disagrees (%)	SE
Netherlands	27,9	7,0	14,9	4,1	57,1	11,5
New Zealand	39,1	3,9	34,3	4,9	26,5	3,4
Norway	10,4	2,5	36,7	6,4	52,9	5,9
Poland	77,5	18,3	4,5	4,6	18,0	10,5
Qatar	59,0	4,3	12,7	2,2	28,4	2,8
Romania	28,0	7,2	39,8	12,4	32,2	7,3
Russian Federation	21,3	2,8	49,4	10,1	29,3	5,7
Scotland	62,4	12,6	28,1	11,7	9,6	3,9
Singapore	23,5	1,2	27,4	1,1	49,1	1,8
Slovakia	42,2	9,5	34,7	14,1	23,1	5,0
Slovenia	30,9	4,9	23,7	4,7	45,4	9,4
Slovenia	30,9	4,9	23,7	4,7	45,4	9,4
South Africa	35,4	1,9	29,1	2,6	35,5	2,2
Spain	29,3	3,4	36,1	4,4	34,6	3,9
Sweden	26,4	3,9	34,8	6,0	38,7	5,3
Taiwan (China)	44,9	4,1	10,7	1,6	44,4	4,3
Trinidad/Tobago	76,8	7,2	4,5	2,2	18,6	3,5

SE: Standard error. Source: PIRLS 2006. Authors own calculations. Calculated using Jackknife methods with the R-package BIFIEsurvey.

Annex B - Table 5: Guardians report on the home language of students who say they speak language of test at home never or sometimes

	Language of test only	SE	Language of test and other language(s)	SE	Only other languages(s)	SE
Australia	23,9	2,6	34,0	3,0	42,1	3,4
Bahrain	22,5	1,2	29,7	1,5	47,8	1,7
Belgium (Flemish)	16,8	1,6	25,3	1,7	58,0	2,2
Bulgaria	14,2	1,9	23,4	2,8	62,4	3,3
Canada	15,0	1,4	32,0	1,7	53,0	2,1
Chile	92,2	1,5	3,4	1,0	4,4	1,2
Croatia	68,9	4,4	18,9	2,9	12,2	3,1
Cyprus	38,3	2,0	22,3	1,3	39,5	1,9
Czech Republic	54,5	3,0	20,3	2,3	25,2	2,4
Denmark	46,4	3,3	15,1	1,7	38,5	3,0
Finland	29,9	3,2	11,1	1,3	59,0	2,8
France	39,5	2,5	38,6	2,0	22,0	2,1
Georgia	56,4	3,8	29,1	2,7	14,5	3,9
Germany	18,4	2,4	24,7	2,5	57,0	2,9
Hong Kong	38,2	2,6	35,7	1,9	26,1	3,1
Hungary	80,7	3,8	5,8	2,2	13,5	3,5
Indonesia	11,2	0,9	15,8	2,0	73,0	2,2
Iran (Islamic Republic)	27,5	2,8	10,4	1,6	62,1	3,1
Ireland	38,0	3,1	20,0	2,2	41,9	3,5
Italy	26,6	2,3	31,3	2,4	42,1	2,2
Japan	82,9	5,8	10,8	4,7	6,3	3,0
Kazakhstan	38,4	2,7	30,0	2,1	31,6	2,8
Korea (Republic)	87,4	1,6	11,4	1,5	1,2	0,6
Kuwait	2,2	0,4	10,0	1,2	87,8	1,3
Lithuania	60,6	3,1	18,4	2,5	21,0	2,6
Morocco	47,8	2,3	15,8	1,2	36,4	2,4
Netherlands	36,5	5,7	34,9	5,8	28,6	6,6
New Zealand	19,1	2,2	37,9	2,5	43,0	2,9



	Language of test only	SE	Language of test and other language(s)	SE	Only other languages(s)	SE
Northern Ireland	26,8	5,9	21,6	4,4	51,5	6,2
Norway	21,3	3,7	22,8	4,9	55,9	4,4
Oman	54,3	1,7	18,5	1,2	27,2	1,7
Poland	80,7	3,3	11,8	2,4	7,5	2,1
Portugal	62,8	3,3	18,0	2,2	19,2	2,7
Qatar	14,0	1,2	29,3	1,3	56,7	1,6
Russian Federation	77,0	9,5	10,4	3,6	12,6	6,2
Saudi Arabia	47,0	3,3	17,3	2,5	35,6	3,3
Serbia	51,0	6,4	23,7	4,7	25,4	7,4
Singapore	3,1	0,3	50,1	0,8	46,8	0,9
Slovakia	36,6	3,5	26,4	2,3	36,9	3,5
Slovenia	53,9	4,4	13,9	2,4	32,1	3,8
Spain	23,6	2,0	13,6	1,2	62,9	2,1
Sweden	16,9	2,2	31,3	2,4	51,8	2,9
Taiwan (China)	15,2	1,1	72,5	1,2	12,3	0,9
Turkey	52,8	3,5	12,1	2,5	35,1	3,7
United Arab Emirates	12,1	0,7	30,2	1,1	57,7	1,2

Source: TIMSS 2015, authors own calculations. Calculated using Jackknife methods with the R-package BIFESurvey.



Annex B - Table 6: Students report on the use of language of instruction at home and on having language of instruction as first language

Country	does not speak / did not learn	SE	does not speak / did learn	SE	speaks / did not learn	SE	speaks / did learn	SE
Austria	5,7	0,6	20,4	0,9	1,9	0,3	71,9	1,2
Belgium (Flemish)	5,6	0,9	17,6	0,9	2,7	0,3	74,2	1,3
Belgium (French)	5,1	0,5	28,8	1,0	2,1	0,3	63,9	1,3
Bulgaria	10,3	1,7	20,0	1,4	2,2	0,3	67,5	2,1
Denmark	3,3	0,4	15,2	0,9	1,2	0,2	80,3	1,1
England	4,6	0,5	19,9	0,9	1,8	0,2	73,7	1,3
France	3,7	0,4	30,8	1,0	1,4	0,2	64,2	1,2
Georgia	3,4	1,6	11,4	0,8	3,0	0,4	82,2	1,5
Germany	4,6	0,5	22,6	0,8	1,4	0,2	71,3	1,0
Hong Kong	2,5	0,3	32,0	0,8	1,5	0,2	64,0	0,8
Hungary	0,8	0,2	24,1	1,1	0,8	0,2	74,3	1,2
Iceland	3,6	0,3	32,5	0,8	3,3	0,3	60,6	0,8
Indonesia	26,0	2,1	35,7	1,7	3,1	0,4	35,2	1,9
Iran (Islamic Republic)	26,0	2,0	21,0	1,5	4,7	0,9	48,3	2,0
Israel	4,0	0,5	38,5	1,1	2,2	0,3	55,3	1,1
Italy	2,4	0,3	26,8	1,1	1,2	0,2	69,6	1,2
Kuwait	49,7	2,3	24,8	1,4	7,3	0,7	18,2	1,3
Latvia	1,3	0,3	29,3	1,4	0,9	0,2	68,5	1,5
Lithuania	1,1	0,2	19,7	1,0	0,7	0,1	78,5	1,0
Luxembourg	66,5	0,6	30,8	0,6	0,7	0,1	2,0	0,2
Macedonia (FYR)	2,4	0,5	14,4	1,0	2,6	0,4	80,6	1,3
Moldova (Republic)	3,3	0,5	22,9	1,0	2,5	0,7	71,3	1,3
Morocco	17,4	2,5	33,0	2,4	4,5	0,8	45,0	2,6
Netherlands	2,3	0,6	22,1	1,1	1,0	0,2	74,6	1,2
New Zealand	6,5	0,5	20,2	0,7	2,0	0,2	71,3	1,0
Norway	3,0	0,4	17,2	0,9	2,9	1,0	76,9	1,2
Poland	0,4	0,1	14,2	0,6	1,3	0,2	84,1	0,6
Qatar	4,8	0,3	34,0	0,6	2,7	0,2	58,4	0,7
Romania	3,3	0,9	15,4	1,1	4,6	1,1	76,8	2,0
Russian Federation	5,0	1,0	13,1	0,8	1,2	0,2	80,7	1,1



Country	does not speak / did not learn	SE	does not speak / did learn	SE	speaks / did not learn	SE	speaks / did learn	SE
Scotland	2,9	0,4	16,9	0,9	2,2	0,3	78,0	1,1
Singapore	15,4	0,5	63,4	0,8	0,5	0,1	20,7	0,7
Slovakia	4,1	1,1	24,6	1,1	1,2	0,2	70,1	1,5
South Africa	12,3	0,9	25,5	0,9	10,2	0,5	51,9	1,2
Spain	11,6	1,0	27,9	0,9	1,7	0,3	58,8	1,4
Sweden	4,8	0,5	20,1	0,9	3,1	0,3	72,0	1,2
Taiwan (China)	4,3	0,4	59,5	0,8	1,3	0,2	34,9	0,8
Trinidad/Tobago	5,6	0,6	16,9	1,0	6,8	1,1	70,6	1,6
USA	5,3	0,5	23,1	1,1	2,3	0,3	69,2	1,3

SE: Standard error. Source: PIRLS 2006, authors own calculations. Calculated using Jackknife methods with the R-package BIFIEsurvey.

ANNEX C: Participation status of countries for different surveys

Country	TIMSS 2019	PIRLS 2016	TIMSS 2015	PIRLS/TIMSS 2011	TIMSS 2007	PIRLS 2006	TIMSS 2003	PIRLS 2001	SACMEQ 4	SACMEQ 3	SACMEQ 2	PASEC 2019	PASEC 2014	LLECE-TERCE 2013	LLECE-SERCE 2006	Some Large Scale Assessment data since 2010	MICS6	Some data (incl. MICS 6)	
Algeria					X												X	X	
Argentina								X						X	X	X	X	X	
Armenia				X	X		X									X		X	
Australia		X	X	X	X		X									X		X	
Austria		X		X	X	X										X		X	
Azerbaijan		X		X												X	X	X	
Bahrain		X	X	X												X		X	
Bangladesh																	X	X	
Belarus																	X	X	
Belize								X											
Benin												X	X			X		X	
Bosnia and Herzegovina																	X	X	
Botswana				X					X	X	X					X		X	
Brazil														X	X	X		X	
Bulgaria		X	X	X		X		X								X		X	
Burkina Faso												X	X			X		X	
Burundi												X	X			X		X	
Cameroon												X	X			X		X	
Canada		X	X	X				X								X		X	
Central African Republic																		X	X
Chad												X	X			X	X	X	
Chile		X	X	X										X	X	X		X	
Colombia				X	X			X						X	X	X		X	
Congo												X	X			X		X	
Congo (DRC)												X				X	X	X	
Costa Rica														X	X	X	X	X	
Cote d'Ivoire												X	X			X		X	



Country	TIMSS 2019	PIRLS 2016	TIMSS 2015	PIRLS/TIMSS 2011	TIMSS 2007	PIRLS 2006	TIMSS 2003	PIRLS 2001	SACMEQ 4	SACMEQ 3	SACMEQ 2	PASEC 2019	PASEC 2014	LLECE-TERCE 2013	LLECE-SERCE 2006	Some Large Scale Assessment data since 2010	MICS6	Some data (incl. MICS 6)
Croatia			X	X												X		X
Cuba															X		X	X
Cyprus			X				X	X								X		X
Czechia		X	X	X	X			X								X		X
Denmark		X	X	X	X	X										X		X
Dominican Republic														X	X	X	X	X
Ecuador														X	X	X		X
El Salvador					X										X		X	X
Equatorial Guinea																	X	X
Finland		X	X	X												X		X
France		X	X	X		X		X								X		X
Gabon												X				X		X
Gambia																	X	X
Georgia		X	X	X	X	X										X	X	X
Germany		X	X	X	X	X		X								X		X
Ghana																	X	X
Greece								X										
Guatemala														X	X	X		X
Guinea												X				X		X
Guinea-Bissau																	X	X
Honduras				X										X		X	X	X
Hungary		X	X	X	X	X	X	X								X		X
Iceland						X		X										
Indonesia			X	X		X										X		X
Iran (Islamic Republic)		X	X	X	X	X	X	X								X		X
Iraq																	X	X
Ireland		X	X	X												X		X
Israel		X		X		X		X								X		X
Italy		X	X	X	X	X	X	X								X		X
Japan			X	X	X		X									X		X



Country	TIMSS 2019	PIRLS 2016	TIMSS 2015	PIRLS/TIMSS 2011	TIMSS 2007	PIRLS 2006	TIMSS 2003	PIRLS 2001	SACMEQ 4	SACMEQ 3	SACMEQ 2	PASEC 2019	PASEC 2014	LLECE-TERCE 2013	LLECE-SERCE 2006	Some Large Scale Assessment data since 2010	MICS6	Some data (incl. MICS 6)
Kazakhstan		X	X	X	X											X	X	X
Kenya									X	X	X					X		X
Korea (DPR)																	X	X
Korea (Republic)			X	X												X		X
Kuwait		X	X	X	X	X		X								X		X
Kyrgyzstan																	X	X
Laos																	X	X
Latvia		X			X	X	X	X								X		X
Lesotho									X	X	X					X	X	X
Lithuania		X	X	X	X	X	X	X								X		X
Luxembourg					X													
Macedonia (FYR)					X		X										X	X
Madagascar												X				X	X	X
Malawi									X	X	X					X		X
Mali												X				X		X
Malta		X		X												X		X
Mauritania																	X	X
Mauritius									X	X	X					X		X
Mexico														X	X	X		X
Moldova (Republic)					X	X	X											
Mongolia				X													X	X
Montenegro																	X	X
Morocco		X	X	X	X	X	X	X								X		X
Mozambique									X	X	X					X		X
Namibia									X	X	X					X		X
Netherlands		X	X	X	X	X	X	X								X		X
New Zealand		X	X	X	X	X	X	X								X		X
Nicaragua														X	X	X		X
Niger												X	X			X		X
Northern Ireland		X	X	X												X		X
Norway		X	X	X	X	X	X	X								X		X



Country	TIMSS 2019	PIRLS 2016	TIMSS 2015	PIRLS/TIMSS 2011	TIMSS 2007	PIRLS 2006	TIMSS 2003	PIRLS 2001	SACMEQ 4	SACMEQ 3	SACMEQ 2	PASEC 2019	PASEC 2014	LLECE-TERCE 2013	LLECE-SERCE 2006	Some Large Scale Assessment data since 2010	MICS6	Some data (incl. MICS 6)
Oman		X	X	X												X		X
Pakistan																	X	X
Panama														X	X	X		X
Paraguay														X	X	X		X
Peru														X	X	X		X
Philippines							X											
Poland		X	X	X		X										X		X
Portugal		X	X	X												X		X
Qatar		X	X	X	X	X										X		X
Romania				X		X		X								X		X
Russian Federation		X	X	X	X	X	X	X								X		X
Sao Tome and Principe																	X	X
Saudi Arabia		X	X	X												X		X
Senegal												X	X			X		X
Serbia			X	X												X	X	X
Seychelles									X	X	X					X		X
Sierra Leone																	X	X
Singapore		X	X	X	X	X	X	X								X		X
Slovakia		X	X	X	X	X		X								X		X
Slovenia			X		X		X	X								X		X
South Africa		X		X		X			X	X	X					X		X
Spain		X	X	X		X										X		X
Sudan																	X	X
Suriname																	X	X
Swaziland									X	X	X					X		X
Sweden		X	X	X	X	X		X								X		X
Tanzania									X	X	X					X		X
Thailand				X												X		X
Togo												X	X			X	X	X
Trinidad and Tobago		X		X		X										X		X



Country	TIMSS 2019	PIRLS 2016	TIMSS 2015	PIRLS/TIMSS 2011	TIMSS 2007	PIRLS 2006	TIMSS 2003	PIRLS 2001	SACMEQ 4	SACMEQ 3	SACMEQ 2	PASEC 2019	PASEC 2014	LLECE-TERCE 2013	LLECE-SERCE 2006	Some Large Scale Assessment data since 2010	MICS6	Some data (incl. MICS 6)
Tunisia				X	X		X									X	X	X
Turkey			X	X				X								X		X
Turkmenistan																	X	X
Uganda									X	X	X					X		X
Ukraine					X												X	X
United Arab Emirates		X	X	X												X		X
United Kingdom																		
United States of America		X	X	X	X	X	X	X								X		X
Uruguay														X	X	X		X
Yemen				X	X		X									X		X
Zambia									X	X						X		X
Zimbabwe									X	X						X	X	X
Count of countries	0	42	43	55	33	33	21	32	14	14	12	15	10	15	16	97	41	12
Other jurisdictions																		
Belgium (Flemish)		X	X	X		X	X									X		X
Belgium (French)		X		X		X										X		X
Buenos Aires, Argentina			X													X		X
England		X	X	X	X	X	X	X								X		X
Hong Kong		X	X	X	X	X	X	X								X		X
Mexican state of Nuevo León														X		X		X
Macao SAR		X														X		X
Ontario, Canada			X	X	X	X	X	X								X		X
Quebec, Canada			X	X	X	X	X	X								X		X
Zanzibar									X	X	X					X		X
Scotland					X	X	X	X										
Taiwan (China)		X	X	X	X	X	X									X		X



Country	TIMSS 2019	PIRLS 2016	TIMSS 2015	PIRLS/TIMSS 2011	TIMSS 2007	PIRLS 2006	TIMSS 2003	PIRLS 2001	SACMEQ 4	SACMEQ 3	SACMEQ 2	PASEC 2019	PASEC 2014	LLECE-TERCE 2013	LLECE-SERCE 2006	Some Large Scale Assessment data since 2010	MICS6	Some data (incl. MICS 6)
	0	42	43	55	33	33	21	32	14	14	12	15	10	15	16	97	41	12
																		3

DRAFT for discussion



6 References

- Albaugh, E. A. (2012). *Language Policies in African States*. Brunswick: Bowdoin College.
- Alidou, H. (2006). *Optimizing Learning and education in Africa - the language Factor: A stock-taking research on Mother Tongue and Bilingual Education in Sub-Saharan Africa*. Paris. Retrieved from http://www.adeanet.org/adea/downloadcenter/Ouga/B3_1_MTBLE_en.pdf
- BIFIE. *BIFIEsurvey: Tools for survey statistics in educational assessment. R package version 2.3-18*. Retrieved from <https://CRAN.R-project.org/package=BIFIEsurvey>
- Breit, S., Bruneforth, M., & Schreiner, C. (Eds.). (2016). *Standardüberprüfung 2015 Deutsch, 4. Schulstufe - Bundesergebnisbericht*. Salzburg. Retrieved from <https://www.bifie.at/node/3360>
- Bruneforth, M., Lassnigg, L., Vogtenhuber, S., Schreiner, C., & Breit, S. (Eds.). (2016). *Nationaler Bildungsbericht Österreich 2015, Band 1: Das Schulsystem im Spiegel von Daten und Indikatoren*. Graz: Leykam.
- Bundesministerium für Bildung. (2016). *Schulformensystematik 2016/17*. Retrieved from https://www.bmb.gv.at/schulen/bw/ueberblick/schulformensystematik_201617.pdf
- Department for Education. (2017). *School census 2017 to 2018: Guide, version 1.4*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/668508/2017_to_2018_School_Census_Guide_V1_4.pdf
- Department of Education and Early Childhood Development. (2013). *SFO Guidelines: Collection of Nationally Consistent Student Background Information*. Retrieved from <http://www.education.vic.gov.au/Documents/school/principals/finance/sfoguidelines.docx>
- Department of Education and Skills. *Primary Online Database (POD)*. Retrieved from <https://www.education.ie/en/Publications/Statistics/Primary-Online-Database-POD/>
- DHS. *DHS Model Questionnaires*. Retrieved from <https://dhsprogram.com/What-We-Do/Survey-Types/DHS-Questionnaires.cfm>
- Elley, W. B. (1992). *How in the world do students read? IEA study of reading literacy*. Hamburg: IEA. Retrieved from <http://eric.ed.gov/?id=ED360613>
- EPDC. *EGRA and EGMA*. Retrieved from <https://www.epdc.org/data-about-epdc-data-epdc-learning-outcomes-data/egra-and-egma>
- Fehrler, S., & Michaelowa, K. (2009). *Education marginalization in Sub-Saharan Africa: Background paper prepared for the Education for All Global Monitoring Report 2010*. Retrieved from <http://unesdoc.unesco.org/images/0018/001865/186584e.pdf>
- Flemish Department of Education and Training. (2015). *OECD REVIEW OF POLICIES TO IMPROVE THE EFFECTIVENESS OF RESOURCE USE IN SCHOOLS. COUNTRY BACKGROUND REPORT OF BELGIUM/FLEMISH COMMUNITY*. Retrieved from http://www.oecd.org/edu/Country%20Background%20Report_Belgium_Flemish%20Community_FINAL%20with%20stat_for%20OECD%20website_REV%20DN.pdf



- Gove, A., & Wetterberg, A. (Eds.). (2011). *The Early Grade Reading Assessment. Applications and Interventions to Improve Basic Literacy*: RTI International. Retrieved from https://www.rti.org/sites/default/files/resources/14019568_bk_0007_1109_wetterberg.pdf
- Grosh, M., & Glewwe, P. (Eds.). (2000). *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study.: Volumes 1, 2, and 3*. Washington.
- Herzog-Punzenberger, B., & Schnell, P. (2012). Die Situation mehrsprachiger Schüler/innen im österreichischen Schulsystem – Problemlagen, Rahmenbedingungen und internationaler Vergleich. In B. Herzog-Punzenberger (Ed.), *Nationaler Bildungsbericht Österreich 2012 - Band 2. Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (pp. 229–268). Graz: Leykam. <https://doi.org/10.17888/nbb2012-2-6>
- Hooper, M., Mullis, I. V.S., & Martin, M. O. (2013). Chapter 3: TIMSS 2015 Context Questionnaire Framework. In I. V.S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks..* Chestnut Hill: TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap3.pdf
- Hooper, M., Mullis, I. V.S., & Martin, M. O. (2015). Chapter 2: PIRLS 2016 Context Questionnaire Framework. In I. V.S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 Assessment Framework*. Chestnut Hill: TIMSS & PIRLS International Study Center.
- Hungi, N. (2011). Characteristics of Grade 6 Pupils, their Homes and Learning Environments. Retrieved from http://www.sacmeq.org/sites/default/files/sacmeq/publications/01_pupils_final_24nov2011.pdf
- IEA (International Association for the Evaluation of Educational Achievement) (Ed.). (2013). *PIRLS 2011. User Guide for the International Database. Supplement 1: International Version of the PIRLS 2011 Background Questionnaires and Curriculum Questionnaire*: TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/pirls2011/downloads/P11_UG_Supplement1.pdf
- International Association for the Evaluation of Educational Achievement (IEA) (Ed.). (2016). *PIRLS 2016. School Questionnaire*: TIMSS & PIRLS International Study Center.
- Kemper, T. (2017). Die schulstatistische Erfassung des Migrationshintergrundes in Deutschland. *Journal for Educational Research Online*, 9(1), 144–168. Retrieved from <http://www.j-e-r-o.com/index.php/jero/article/viewFile/737/308>
- Kosonen, K. (2017). Language of instruction in Southeast Asia: Background paper prepared for the Global Education Monitoring Report 2017/8. Retrieved from <http://unesdoc.unesco.org/images/0025/002595/259576e.pdf>
- LaRoche, S., & Foy, P. (2016). CHAPTER 5: Sample Implementation in TIMSS 2015. In M. O. Martin, I. V.S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015*. (5.1-5.175). Chestnut Hill.
- Leadership Council of the Sustainable Development Solutions Network. (2015). Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for



- the SDGs. Retrieved from <http://unsdsn.org/wp-content/uploads/2015/05/150612-FINAL-SDSN-Indicator-Report1.pdf>
- LLECE. (2010). Compendio de los Manuales del SERCE. Retrieved from <http://unesdoc.unesco.org/images/0019/001919/191940s.pdf>
- Martin, M. O., Mullis, I. V.S., & Foy, P. (2015). Chapter 3: Assessment Design for PIRLS, PIRLS Literacy, and ePIRLS in 2016. In I. V.S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 Assessment Framework* (pp. 55–70). Chestnut Hill: TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/pirls2016/downloads/P16_Framework_2ndEd.pdf
- Mellander, E. (2016). *On the use of register data in educational science research. IFAU WORKING PAPER*. Retrieved from <http://www.ifau.se/globalassets/pdf/se/2016/wp2016-22-on-the-use-of-register-data-in-educational-science-research.pdf>
- Ministry of Education Ghana. (2016). Ghana 2016 National Education Assessment: Report of Findings. Retrieved from http://pdf.usaid.gov/pdf_docs/PA00MHMR.pdf
- Mullis, I. V.S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 Assessment Framework*. Chestnut Hill: TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/pirls2016/downloads/P16_Framework_2ndEd.pdf
- Mullis, I. V.S., Martin, M. O., Goh, S., & Cotter, K. TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. Retrieved from <http://timssandpirls.bc.edu/timss2015/encyclopedia/>
- Mullis, I. V.S., Martin, M. O., Kennedy, A., & Foy, P. (Eds.). (2007). *PIRLS 2006 International Report*: TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/PDF/PIRLS2006_international_report.pdf
- Nusche, D. (2009). *What Works in Migrant Education? OECD Education Working Papers: No. 22*. Paris: OECD Publishing.
- OECD (Ed.). (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment. OECD Reviews of Evaluation and Assessment in Education*. Paris: OECD Publishing.
- Ouane, A., & Glanz, C. (2011). *Optimising Learning, Education and Publishing in Africa: The Language Factor A Review and Analysis of Theory and Practice in Mother-Tongue and Bilingual Education in sub-Saharan Africa*. Hamburg/Abidjan: UNESCO Institute for Lifelong Learning; Association for the Development of Education in Africa. Retrieved from <http://unesdoc.unesco.org/images/0021/002126/212602e.pdf>
- Pinnock, H. (2009). *Language and education: the missing link: How the language used in schools threatens the achievement of Education For All*. Retrieved from <http://www.unesco.org/education/EFAWG2009/LanguageEducation.pdf>
- Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (Ed.). (2015). *PASEC 2014: Education System Performance in Francophone Sub-Saharan Africa*. Dakar. Retrieved from http://www.pasec.confemen.org/wp-content/uploads/2015/12/Rapport_Pasec2014_GB_webv2.pdf



- RTI International. (2009). Early Grade Reading Assessment Toolkit. Retrieved from http://s3.amazonaws.com/inee-assets/resources/EGRA_Toolkit_Mar09.pdf
- RTI International. (2014a). Research on Reading in Morocco: Analysis of Initial Teacher Training Final Report: Component 2: Data for Education Programming in Asia and the Middle East (DEP/AME). Retrieved from https://globalreadingnetwork.net/sites/default/files/eddata/R1-010b_Morocco_Reading_Study_2-ENG-Translated.pdf
- RTI International. (2014b). Tout Timoun Ap Li – ToTAL: Year 2 EGRA and SSME Endline Report, REVISED. Retrieved from https://globalreadingnetwork.net/sites/default/files/eddata/TechR_RTI_Haiti_ToTAL_Task_18_Y2_EGRA-SSME_Endline_Report_REVISED_w_AnnexA.pdf
- RTI International. (2015). *Early learning assessments: A retrospective. (Background paper for EFA Global Monitoring Report 2015.)*. Retrieved from <unesdoc.unesco.org/images/0023/002324/232419e.pdf>
- SACMEQ. "Part B: Sample Design Procedures for the SACMEQ II Project". Retrieved from <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/226/download/3576>
- SACMEQ. (2000). SACMEQ II Questionnaires. Retrieved from <http://microdata.worldbank.org/index.php/catalog/1245>
- SACMEQ. (2007). SACMEQ III Questionnaires. Retrieved from <http://microdata.worldbank.org/index.php/catalog/1246/download/22687>
- Salchegger, S., & Suchań, B. (2017). Was bedeutet es für den Geschlechterunterschied in der Mathematikkompetenz bei PISA, wenn dem Schulsystem leistungsschwache Jungen verloren gehen? *Zeitschrift für Bildungsforschung*, 135(7). <https://doi.org/10.1007/s35834-017-0190-7>
- Schulte, K., Hartig, J., & Pietsch, M. (2016). Der Sozialindex für Hamburger Schulen. In B. Groot-Wilken, K. Isaac, & J.-P. Schräpler (Eds.), *Socialindices für Schulen. Hintergründe, Methoden und Anwendung* (pp. 67–80). Waxmann. Retrieved from <http://www.hamburg.de/contentblob/4458462/5f7d1b2af92c58e7328be298374bbb05/data/pdf-schulte-und-hartig.pdf>
- Scottish Government (ScotXed). (2017). Survey Documentation. Retrieved from <http://www.gov.scot/Topics/Statistics/ScotXed/SchoolEducation/SchoolPupilCensus/SurveyDocumentation>
- Smits, J., & Gündüz-Hoşgör, A. (2003). Linguistic capital: Language as a socio-economic resource among Kurdish and Arabic women in Turkey. *Ethnic and Racial Studies*, 26(5), 829–853. <https://doi.org/10.1080/0141987032000109050>
- Sprengel, L. (2008). Senegal: Early Grade Reading Assessment (EGRA). Retrieved from <https://openknowledge.worldbank.org/bitstream/handle/10986/12974/697150ESW0P1050ive0Senegal00English.pdf>
- Suchań, B., & Schreiner, C. (Eds.). (2012). *PIRLS & TIMSS 2011. Technischer Bericht.: (Österreich)*. Salzburg. Retrieved from https://www.bifie.at/wp-content/uploads/2017/10/PIRLS_TIMSS_2011_Technischer-Bericht.pdf



- UNESCO (Ed.). (2016). *If You Don't Understand, How Can You Learn? Global Education Monitoring Report Policy Paper: Vol. 24*. Paris: Unesco Publishing. Retrieved from <http://unesdoc.unesco.org/images/0024/002437/243713E.pdf>
- UNESCO (Ed.). (2017). *Global Education Monitoring Report 2017/8: Accountability in education: Meeting our Commitments*. Paris: Unesco Publishing. Retrieved from <http://unesdoc.unesco.org/images/0025/002593/259338e.pdf>
- UNESCO Bangkok. (2008). *Using a literacy module in household surveys: A guidebook*. Bangkok. Retrieved from http://uis.unesco.org/sites/default/files/documents/using-a-literacy-module-in-household-surveys-a-guidebook-en_0.pdf
- UNESCO IBE, & UNESCO Institute for Statistics. Monitoring Progress towards SDG 4.1: Initial Analysis of National Assessment Frameworks for Mathematics. Retrieved from <http://unesdoc.unesco.org/images/0025/002596/259685E.pdf>
- UNESCO Institute for Statistics. (2017a). Metadata for the global and thematic indicators for the follow-up and review of SDG 4 and Education 2030. Retrieved from http://uis.unesco.org/sites/default/files/documents/metadata-global-thematic-indicators-sdg4-education2030-2017-en_1.pdf
- UNESCO Institute for Statistics. (2017b). *UIS Indicator Development in the Field of Education. Information Paper: No. 45*. Montreal.
- UNESCO. (2003). *Education in a Multilingual World: UNESCO Education Position Paper*. Retrieved from <http://unesdoc.unesco.org/images/0012/001297/129728e.pdf>
- UNICEF-MICS (Ed.). (2017). *MICS 6: Questionnaire for Children Age 5-17*. New York. Retrieved from <http://mics.unicef.org/tools#survey-design>
- United Nations. (2008). *Principles and recommendations for population and housing censuses (Rev. 2). United Nations publication: Vol. 67*. New York: United Nations. Retrieved from https://unstats.un.org/unsd/publication/SeriesM/Seriesm_67rev2e.pdf
- Wasanga, P. M., Ogle, M. A., & Wambua, R. M. (2012). *The SACMEQ III Project in Kenya - A study of the conditions of schooling and the quality of education*. Retrieved from http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/national-reports/kenya_national_report.pdf
- WorldBank. (2015). MYANMAR EARLY GRADE READING ASSESSMENT FOR THE YANGON REGION. Retrieved from <https://openknowledge.worldbank.org/bitstream/handle/10986/26957/ACS13261-REVISED-PUBLIC-Myanmar-Early-Grade-Reading-Assessment-IDU.pdf>



DRAFT for discussion