

The contribution of student achievement tests to the monitoring of SDGs

Nadir Altinok, BETA, CNRS and University of Lorraine

Incomplete version – Please do not quote

Introduction

- SDG4: Education Goal which aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all".
- Among all indicators provided, the international community has to make critical measurement challenges on two main group of indicators: learning outcomes and educational equality.
- Currently, there is lack of consistency involving standards and definitions between all stakeholders and even between the international organizations which are involved in the production of education statistics (UIS, 2016).

Introduction

- **Five of the ten** education targets focus on the learning outcomes of young children, youth and adults.
- This is a clear shift from MDGs which mainly focused on access, participation and completion.
- The SDG agenda, beyond Goal 4 (and indicator 4.5), highlights the need to focus on equity.
- Hence, education indicators should not only capture national averages but also the variation across different sections of the population defined by group and individual characteristics, such as sex, wealth, location, ethnicity, language or disability.

Introduction

- Measuring learning is complex.
- Despite the growing number of learning assessments, there is currently no framework to put together the various types of assessments and to produce cross-nationally comparable data.
- Target 4.1 covers the quality of primary and lower secondary education.
- The current global indicator for this target is the "proportion of children and young people: (i) in Grade 2 or 3; (ii) at the end of primary education; and (iii) at the end of lower secondary education who achieved at least a minimum proficiency level in (a) reading and (b) mathematics".
- In this study, we will focus on cross-national initiatives (international and regional student achievement tests, ISATs & RSATs).

Introduction

- One of the main challenges for measurement on the global level relates to the definition of what counts as meeting a 'minimum proficiency level' (MPL) in different national contexts, and thus to generate tools to describe the level of proficiency.
- In this study, we propose to prepare a **dataset focused on learning skills**, for measuring both minimum proficiency levels and equity issues.
- Firstly, we will analyze the definition to give to the minimum levels of competence and performance levels.
- Moreover, based on the collection of background questionnaires, our contribution aims at collecting contextual data on education systems, which is often lacking for most countries.

Assessments used in this study

- We selected the latest available assessments
 - International student achievement tests (ISATs)
 - PISA 2012
 - TIMSS 2011
 - PIRLS 2011
 - Regional student achievement tests (RSATs)
 - TERCE 2013
 - SACMEQ III
 - PASEC 2014

2. Student Learning Assessments



TIMSS & PIRLS

- Grades 4 & 8, latest data available for 2011
- MPL = « Low International Benchmark » [400 points]



PISA

- 15 years old students, latest data available for 2012
- MPL = Level 2 for mathematics [407 points] and reading [420 points]



TERCE

- Grades 3 & 6, latest data available for 2013
- MPL = Level 2 for mathematics (G3: 688 / G6: 686 points) and reading (G3: 676 / G6: 613 points)



SACMEQ

- Grade 6, latest data available for 2010
- MPL = Level 3 « Basic Numeracy » for mathematics [466 points] and reading « Basic Reading » [414 points]



PASEC

- Grades 2 & 6, latest data available for 2014
- MPL = Level 2 for mathematics (G2: 400.3 / G6: 433.3 points) and reading (G2: 469.5 / G6: 441.7 points)

2. Student Learning Assessments

International Student Achievement Tests (ISATs)

PISA

Every 3 years since 2000 – 15 years old students
- Latest Year: 2012 – 65 countries

TIMSS

Every 4 years since 1995 – Grades 4 & 8 - Latest
Year: 2011 – 57 countries

PIRLS

Every 5 years since 2001 – Grade 4 – Latest
Year: 2011 – 57 countries

Regional Student Achievement Tests (ISATs)

TERCE

3 rounds since 1996 – Grades 3 & 6 – Latest
Year: 2013 – 15 countries

SACMEQ

4 rounds since 1995 – Grade 6 – Latest Year:
2010 – 15 countries

PASEC

Every 5 years since 2014 – Grades 2 & 6 – Latest
Year: 2014 – 10 countries

The need of obtaining a first overview of the MPLs

- Currently, there is no common metric which may be able to make comparable all international and regional achievement tests.
- Possibility to look at official **minimum proficiency levels** (MPLs) in order to obtain some data about the proportion of students having at least basic skills.
- Skills used for tracking SDGs are mathematics and reading.
- Three education levels: lower primary education, upper primary education and lower secondary education.
- All ISATs and RSATs provide specific benchmarks and a minimum threshold.

The need of obtaining a first overview of the MPLs

- 20 different combinations possible:
 - 5 assessments: PISA, TIMSS, PIRLS, TERCE, SACMEQ and PASEC
 - 5 Grades: Grade 2, Grade 3, Grade 4, Grade 6 & Grade 8
 - 3 skills: Mathematics, Reading & Science (+Environment science & Geoscience)
- The definition of MPLs are taken from official reports or specific documents:
 - PISA/TERCE/PASEC(maths) : LEVEL 1
 - TIMSS/PIRLS: Low International Benchmark
 - SACMEQ/PASEC(reading): LEVEL 2

1. Introduction

2. Data

3. Methodology

4. Results

5. Limits

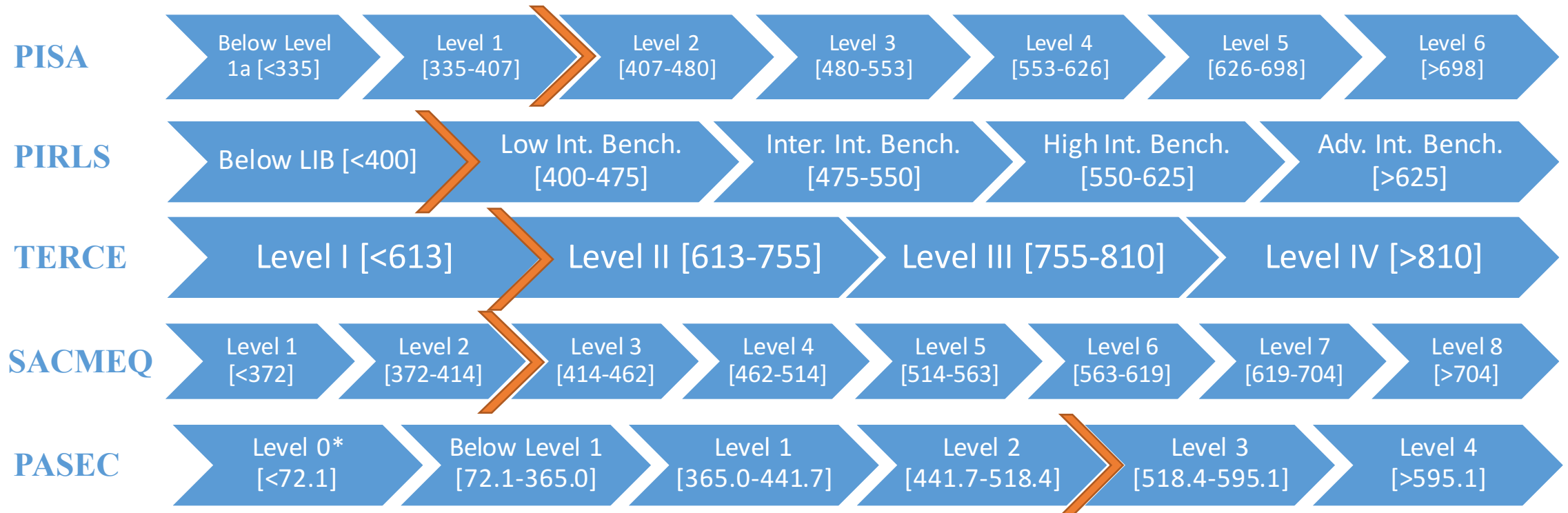
6. Conclusion

Minimum proficiency levels (MPL) for mathematics – Grades 6-8



* In PASEC, no specific name was given to the level below the threshold of 68.1 points. We called it « Level 0 »

Minimum proficiency levels (MPL) for reading – Grades 4-8



* In PASEC, no specific name was given to the level below the threshold of 72.1 points. We called it « Level 0 »

Equity dimension

	Gender	Location	SES	Language	Immigrant	Disability
PISA	✓	✓	✓	✓	✓	
TIMSS	✓	✓	✓	✓	✓	
PIRLS	✓	✓		✓	✓	
TERCE	✓	✓	✓	✓	✓	
SACMEQ	✓	✓	✓	✓		
PASEC	✓	✓	✓	✓		

- Indicator 4.5.1. : Parity indices for all education indicators
- Gender, location and language are available in all assessments
- Disability is lacking from all assessments
- Measures of socio-economic status are included in IEA assessments, but are only focused on home resources.

Equity dimension

- In all ISATs and RSATs, equity dimensions are available.
- However, apart the case of gender, questions asked are not always similar.
- Wealth indicators differ greatly among assessments and are scaled between specific extreme values.
- Location of schools may provide useful information regarding the distinction between urban and rural areas.
- The questions asked differ greatly and depend on the economic level of countries
 - In PASEC and SACMEQ, the definition of a rural area is greatly different from the remaining assessments.
 - Possibility to use the prepared variables for urban/rural areas from ISATs and RSATs.

Example of equity variable: Location of Schools

- In **PISA**, 5 different possibilities for location of schools, while 6 possible answers are available in **TIMSS**:

Q Which of the following definitions best describes the community in which your school is located?

(Please tick only one box.)

A village, hamlet or rural area (fewer than 3 000 people) ₁

A small town (3 000 to about 15 000 people) ₂

A town (15 000 to about 100 000 people) ₃

A city (100 000 to about 1 000 000 people) ₄

A large city (with over 1 000 000 people) ₅

5

A. How many people live in the city, town, or area where your school is located?

*Check **one** circle only.*

More than 500,000 people --

100,001 to 500,000 people --

50,001 to 100,000 people --

15,001 to 50,000 people --

3,001 to 15,000 people --

3,000 people or fewer --

Example of equity variable: Location of Schools

- In **TERCE**, 5 different possibilities for location of schools, while the definition of rural areas are quite relative for **PASEC** and **SACMEQ**:

11. Su escuela se encuentra en una localidad de:

Marque con una X solo una opción.

11.1 2.000 habitantes o menos.	<input type="checkbox"/>
11.2 Entre 2.001 y 5.000 habitantes.	<input type="checkbox"/>
11.3 Entre 5.001 y 10.000 habitantes.	<input type="checkbox"/>
11.4 Entre 10.001 y 100.000 habitantes.	<input type="checkbox"/>
11.5 Más de 100.000 habitantes.	<input type="checkbox"/>

24. Votre école est située dans... ?

(Veuillez ne cocher qu'une seule case)

- Une ville
- Une banlieue de grande ville
- Un grand village (plusieurs centaines de concessions)
- Un petit village (plusieurs dizaines de concessions)

Availability of additional indicators

- ISATs and RSATs not only provide data for the proportion of students reaching the MPLs, but also additional information about the structure of education systems like the availability of electricity in schools.
- Definition and information about the 43 indicators are provided from the TAG Report: “Thematic Indicators to Monitor the Education 2030 Agenda” (October 2015). [link](#)
- Example of indicator n°10: “Participation rate in early childhood care and education in a given period prior to entry into primary education” (indicator 4.2.2)
- Alternative data to the one provided by the UIS are available from ISATs and RSATs, although these data are not fully comparable.
- While in some assessments, we can only know if a student took part to some pre-primary education, in others like TERCE, we can exactly know how many years was this participation.

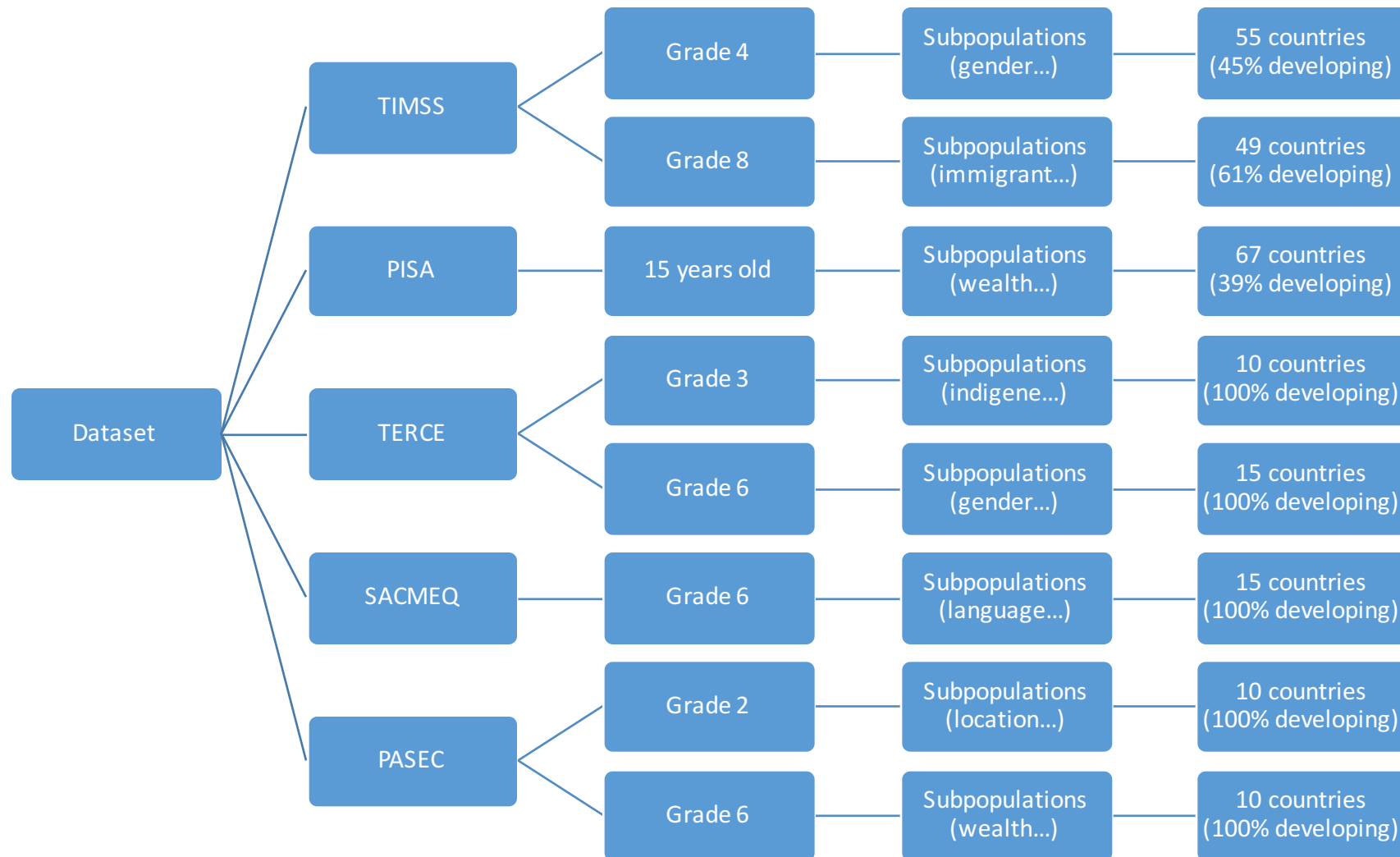
Availability of additional indicators

	PISA	TIMSS	PIRLS	TERCE	SACMEQ	PASEC
Indicator n°6	✓	✓	✓	✓	✓	✓
Indicator n°8			✓			
Indicator n°9			✓			
Indicator n°10	✓	✓	✓	✓	✓	✓
Indicator n°18	✓	✓	✓	✓	✓	✓
Indicator n°27	✓					
Indicator n°28		✓			✓	
Indicator n°30				✓	✓	✓
Indicator n°31				✓	✓	✓
Indicator n°33	✓	✓	✓	✓	✓	✓
Indicator n°37	✓	✓	✓	✓	✓	✓
Indicator n°39				✓	✓	✓

Structuration of the database

- Version 1 of the **international database on minimum proficiency levels**
 - Data for 126 countries or localities
 - Approximately 62% of developing countries.
 - Overall mean of proportion of student reaching the MPL: **71%**
 - Data available for each assessment, grade, skill and subpopulation.
 - Subpopulations: gender, location, wealth, immigration status, language at home and indigene populations (Latin American countries).
 - Approximately **5,000** different combinations of data relative to proportion of student reaching the MPLs.
 - Only latest available assessments: TIMSS 2011, PIRLS 2011, PASEC 2014, PISA 2012, SACMEQ III, TERCE 2013
 - Definitions: Lower Primary = Grades 2 & 3 ; Upper Primary = Grades 4 & 6 ; Lower Secondary = Grades 7-9.

Structuration of the database



Descriptive statistics: low data availability for lower secondary education

- Data for lower primary education is available only for 25 developing countries (based on TERCE and PASEC assessments).
- Upper primary education is the most tested level across ISATs and RSATs and data are available for about 64 developing countries.
- Data is still lacking for the lower secondary education in Sub-Saharan African countries, although data is available for about 45 developing countries.
- If we aggregate all levels, skills, years and assessments, the lowest performing region is **Northern Africa**, followed by **Sub-Saharan Africa** and Western Asia.
- Sub-Saharan African countries perform very poorly in lower secondary (note: very few SSA countries with data for this level)

Descriptive statistics: low data availability for lower secondary education

Table 1. Descriptive statistics on minimum performers, by level of education

Level	Countries	% Developing countries	Mean	Standard Deviation	Minimum	Maximum
Lower Primary	25	100	50.29	21.49	09.60	96.70
Upper Primary	101	64	75.35	22.97	06.66	99.64
Lower Secondary	90	51	67.14	20.47	14.50	98.90
Total	126	62	68.92	22.37	06.66	99.64

Note: The minimum proficiency levels are not directly comparable between assessments. A direct comparison should be made with caution. Lower primary includes grades 2 & 3, Upper primary includes grades 4, 5 & 6, Lower secondary includes grades 7, 8 & 9.

Figure 1. Proportion of students reaching the MPL
(Mean for all levels, grades, skills and assessments)

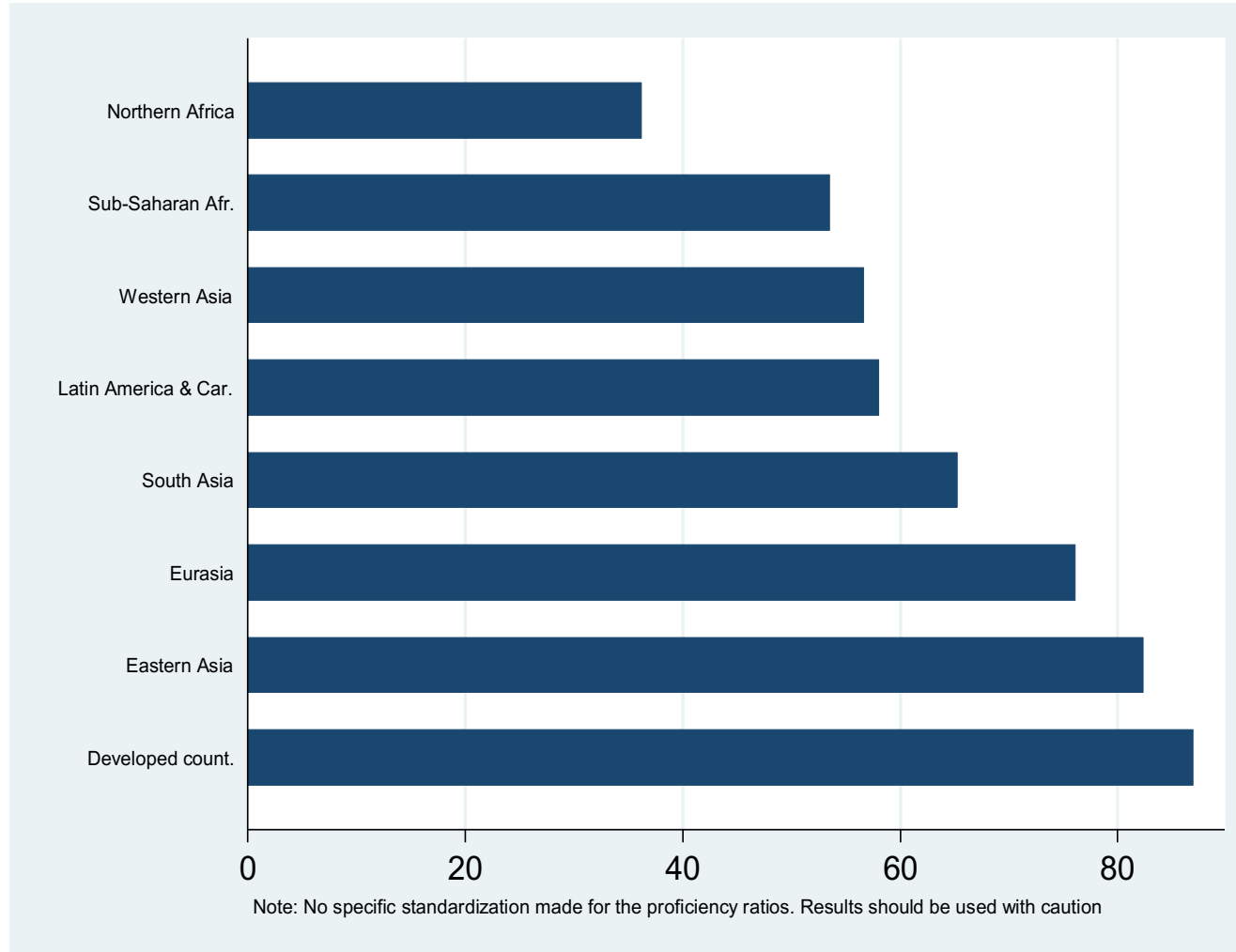
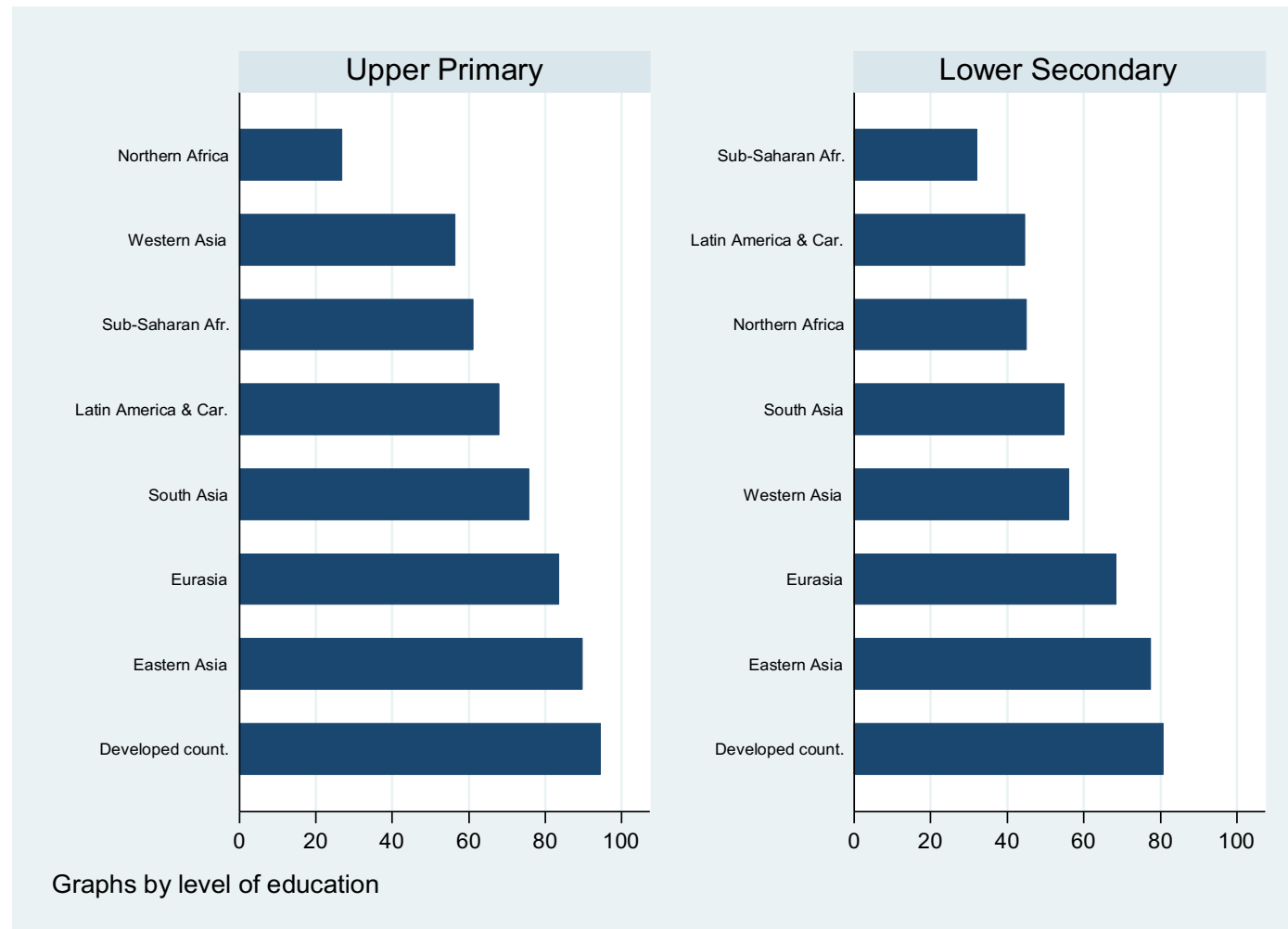


Figure 2. Proportion of students reaching the MPL by education level
(Mean for all skills and assessments)



Descriptive statistics: possibility to combine completion and learning outcomes

- Since we have data for proportion of students reaching the MPLs, we correlated it with survival rates to the last grade of each level.
- Correlation is positive but very low in both levels: only 30% in primary education and less than 10% in lower secondary level.
- While the survival rate in primary education is about the same in Niger and Senegal, the proportion of students reaching the MPLs differs greatly (10% versus 60% respectively, Figure 3.1).
- In lower secondary education, a similar comparison between Peru and Vietnam can be made in favor of the latter country (Figure 3.2).
- When we multiply both indicators, we can obtain a **quality-adjusted survival rate to the last grade** of each education level (figures 4.1. & 4.2.).

Figure 3.1. Relationship between proportion of pupils reaching the MPL and survival rate to last grade, primary education

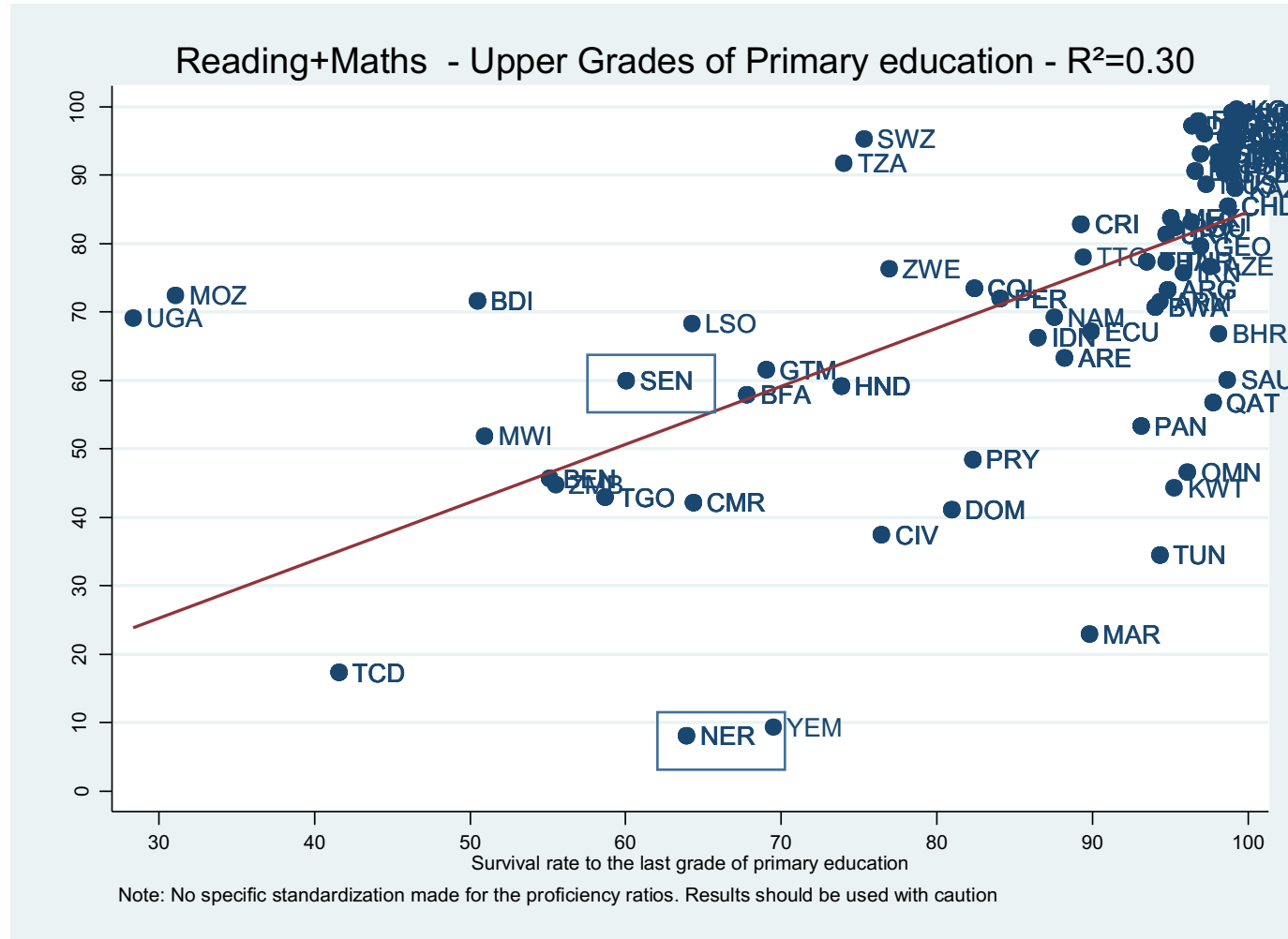


Figure 3.1. Relationship between proportion of pupils reaching the MPL and survival rate to last grade, lower secondary education

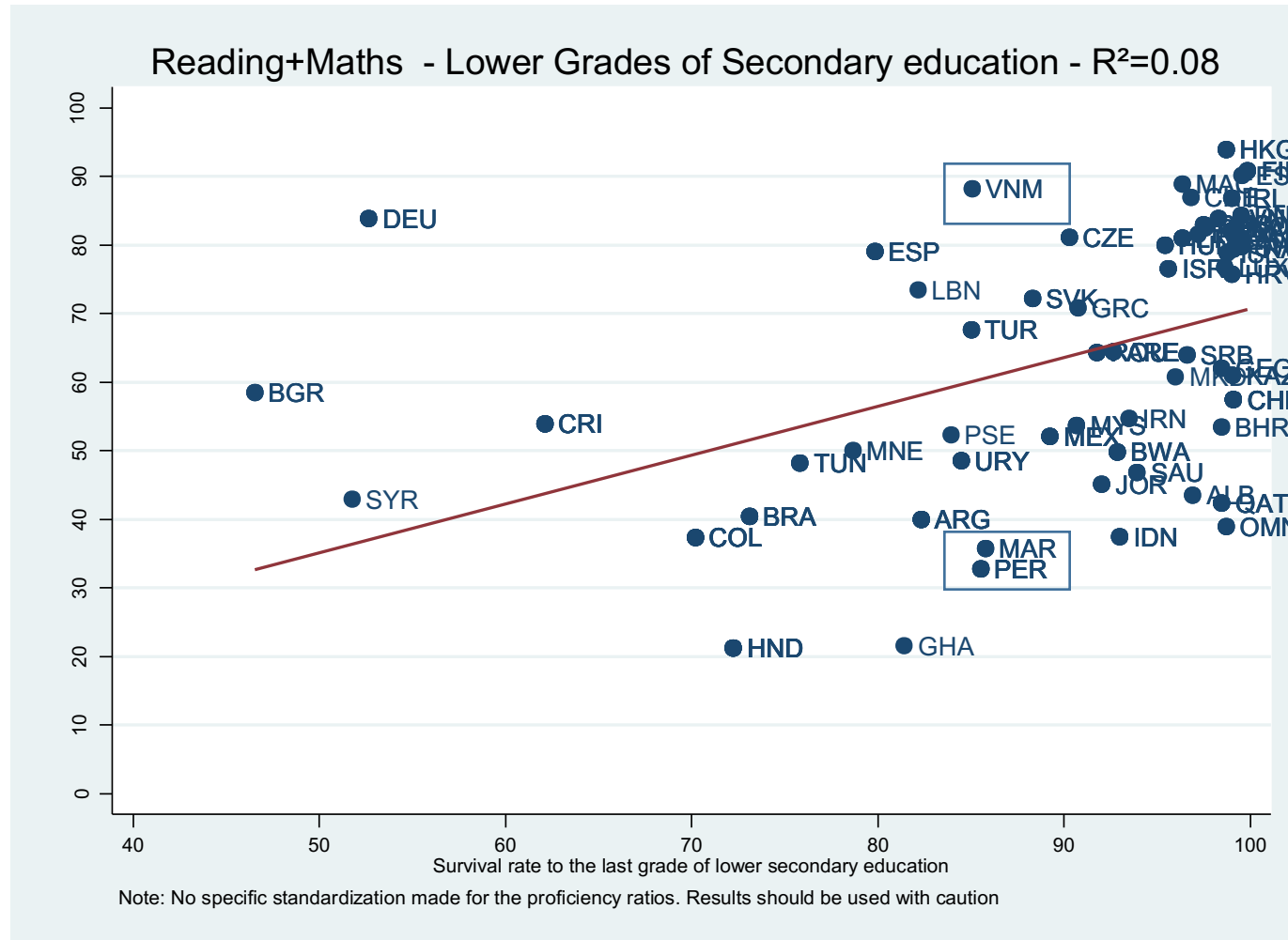


Figure 4.1. Distribution of the quality adjusted survival rate to the last grade, Primary education, Sub-Saharan African countries

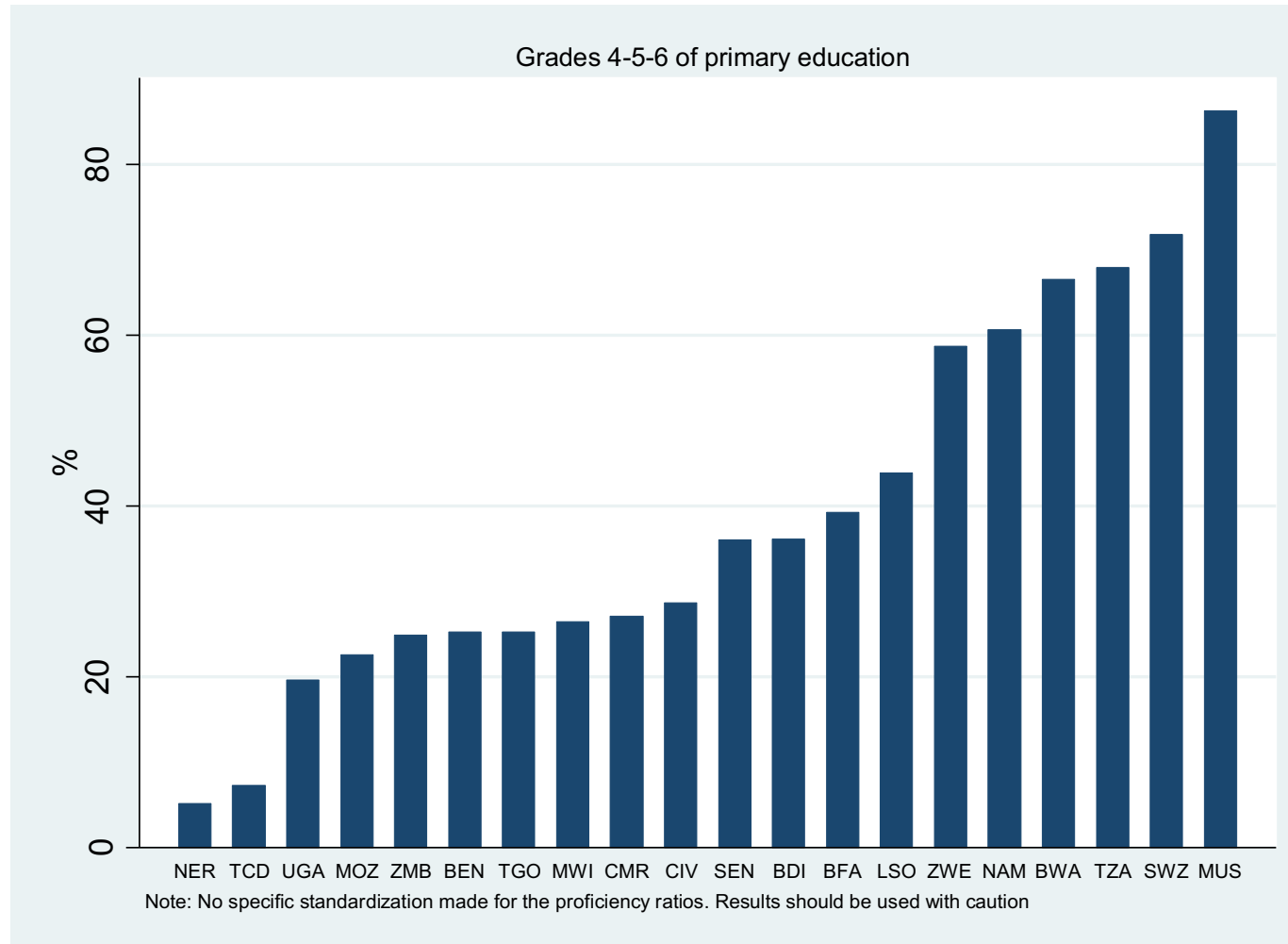
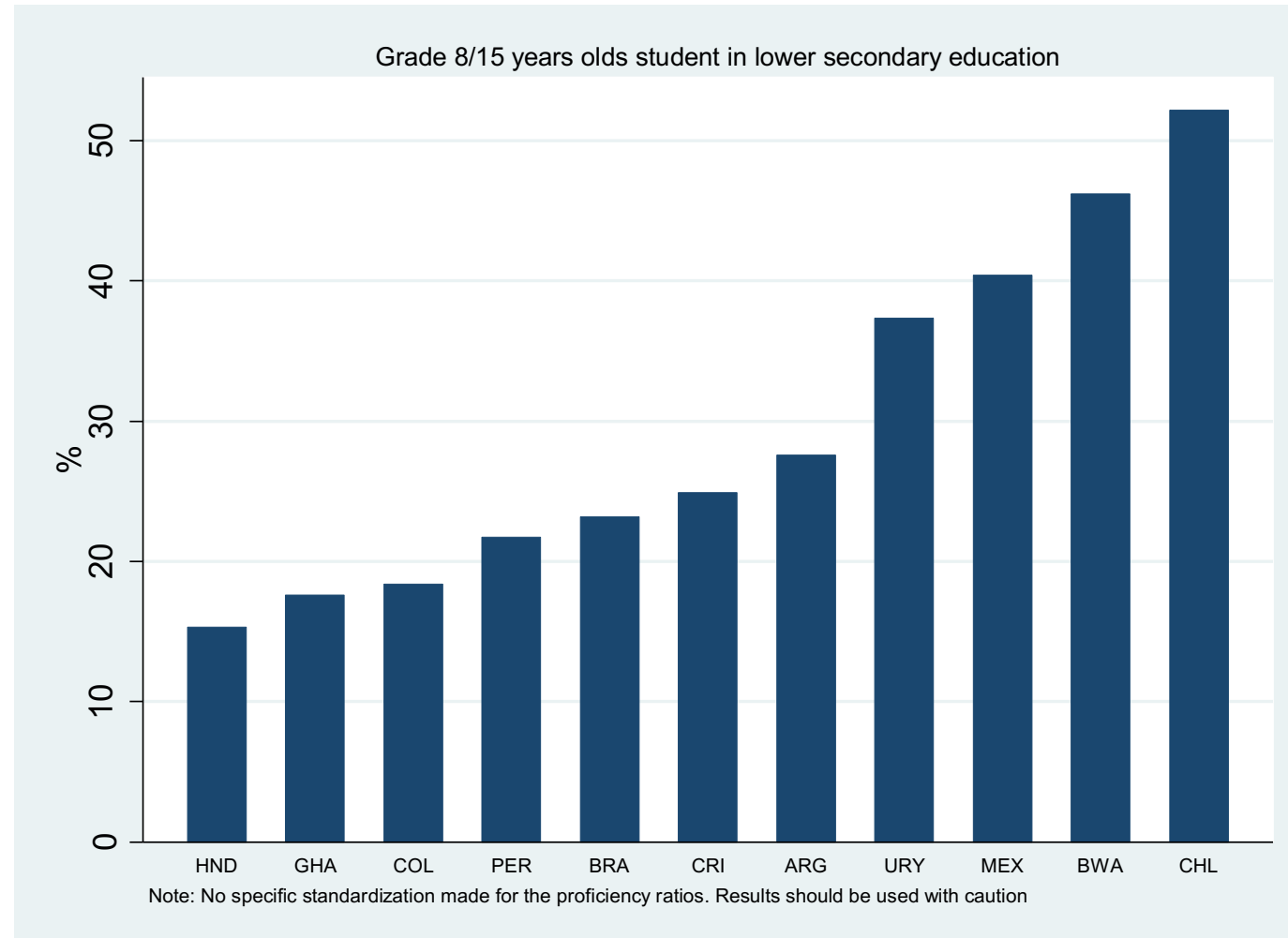


Figure 4.2. Distribution of the quality adjusted survival rate to the last grade, Lower Secondary education, SSA & Latin American countries



4.2. Results for equity issues

- As expected a positive and significant relationship is found between gender parity index (GPI) of students reaching the MPLs and the proportion of students reaching the MPLs.
- Gender Parity Index = Proportion of girls/Proportion of boys reaching the MPLs (when girls perform better, the relationship is reversed)
- If a given country is able to enroll most students and these one are able to learn the minimum learning skills, differences between subpopulations may be lower (i.e. positive relationship)
- It is also possible that some countries perform highly but fail to reach an equitable situation between girls and boys
 - Comparison between **Oman and Benin**
 - Note: comparability issues are less possible regarding the equity issues, since we use ratios.
- Results for Location Parity Index (LPI) and Wealth Parity Index (WPI) highlight a higher inequality between subpopulations (Figure 7).
 - Difference between Peru and Ecuador for location parity index

Figure 5. Minimum Learning and Gender Parity Index, Upper Primary Education

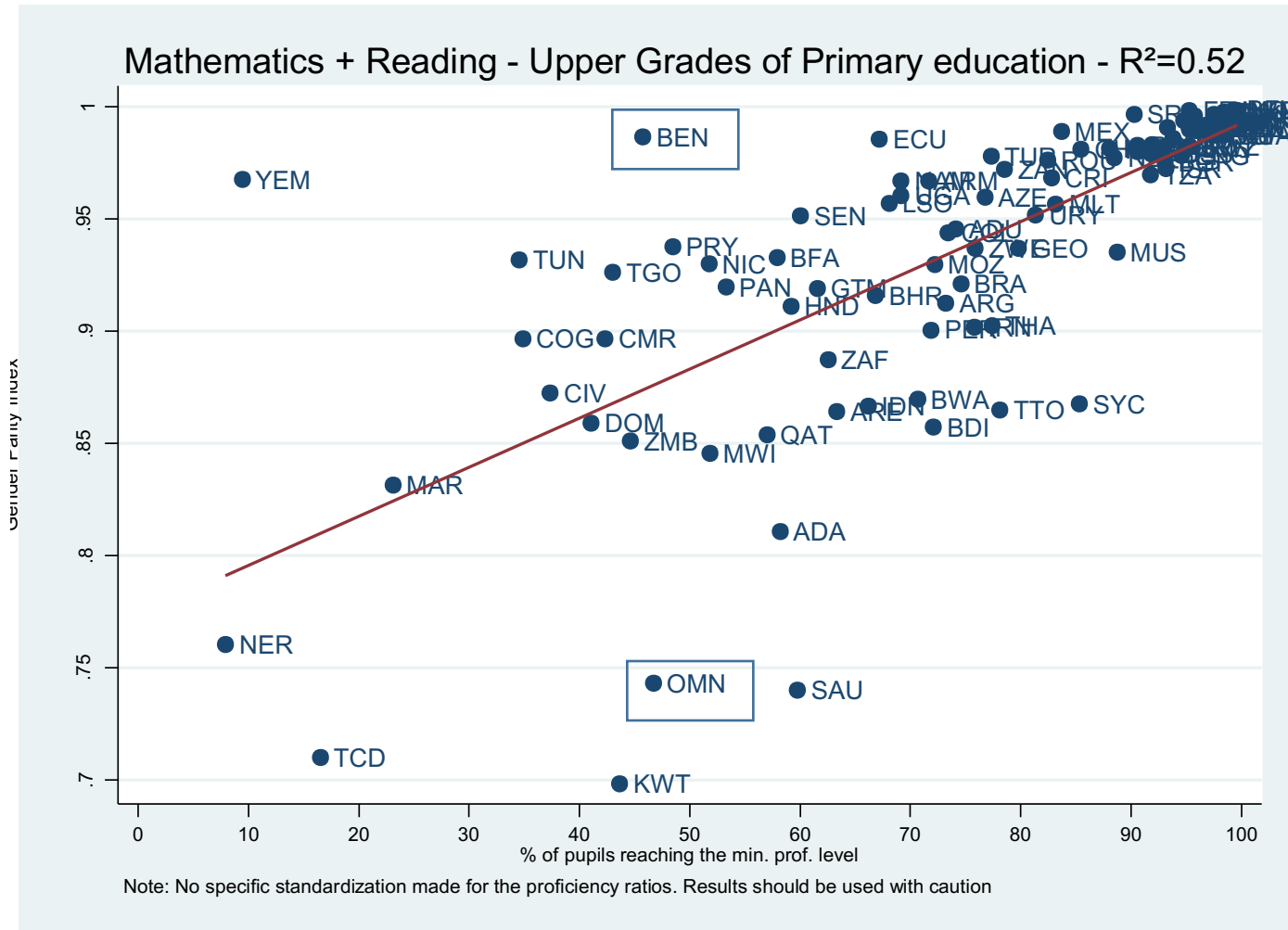
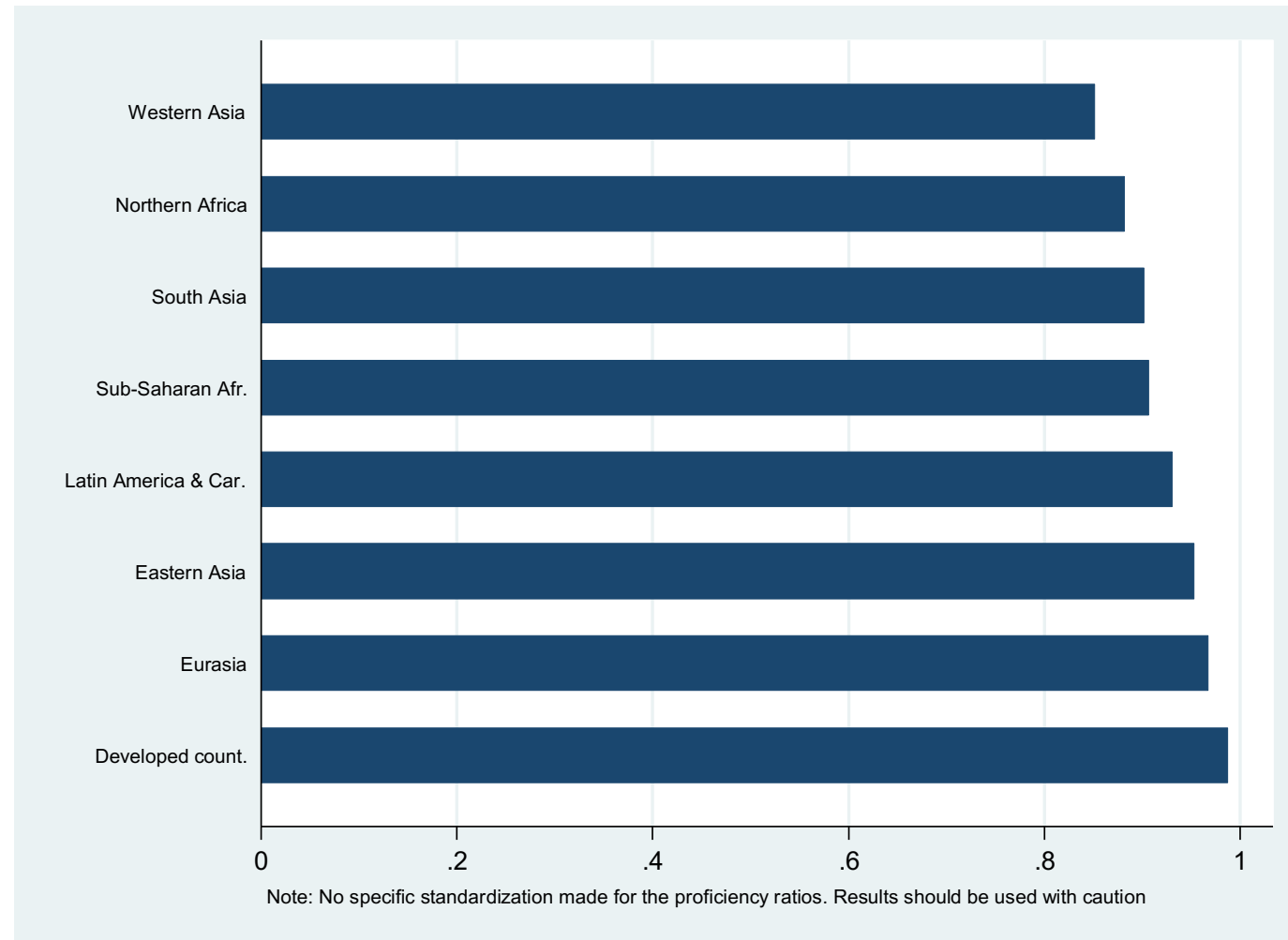


Figure 6. Minimum Learning and Gender Parity Index, Upper Primary Education, by regions



4.3. Results for others indicators

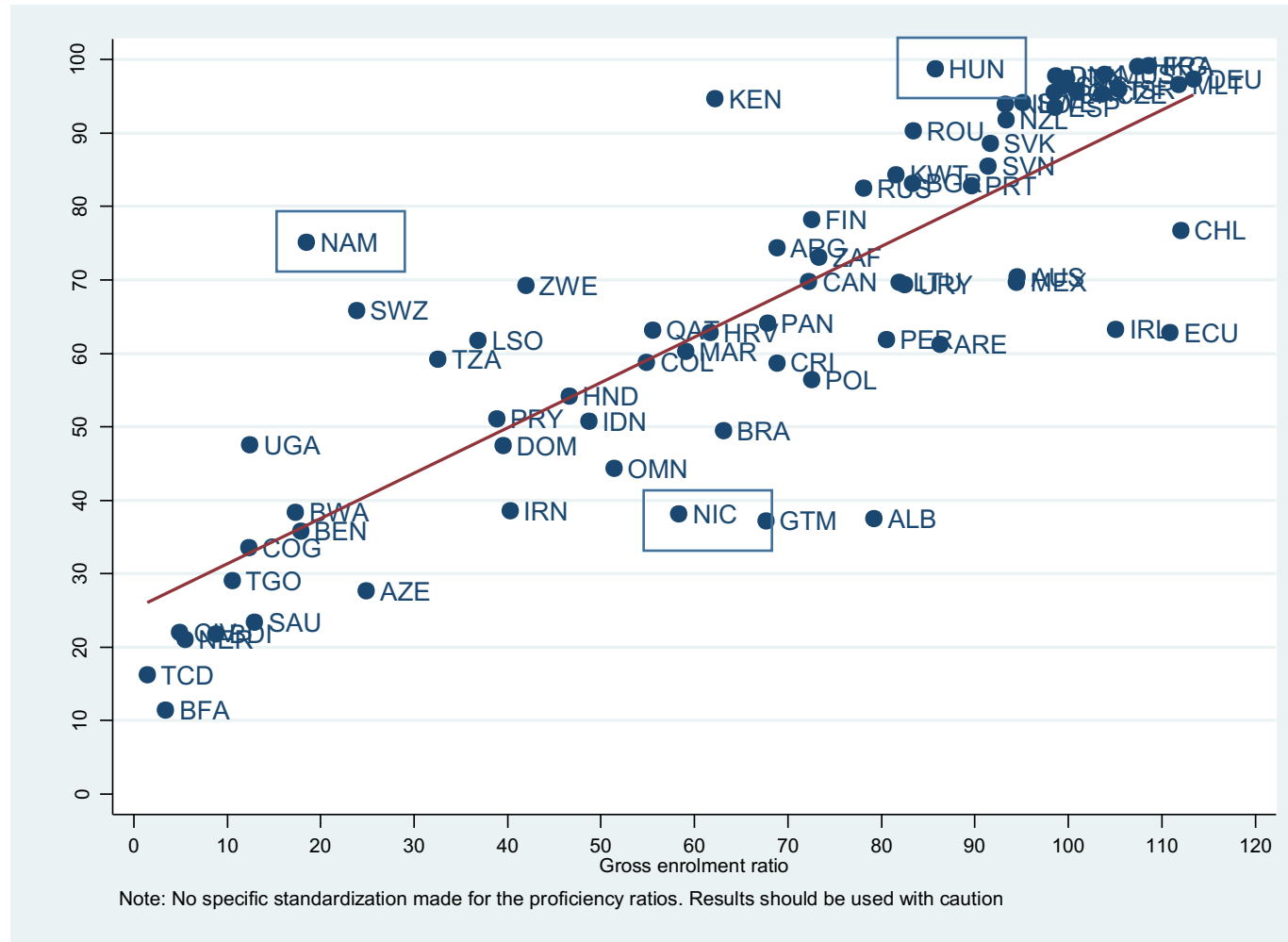
- Alternative indicators may be obtained from ISATs and RSATs.
- Definitions and the choice of indicators taken from the TAG Report published in October 2015.
- Questions asked may not always fit in order to obtain cross-country comparable data.
- The dataset obtained may however be a tool in order to adjust upcoming assessments and obtain standardized definitions.
- For instance, data for pre-primary education can be extracted from ISATs and RSATs.
 - Indicator 10 (SDG Indicator 4.3.1) : « Percentage of children who attended one year of formal education before entering in primary education »
 - Results are not exactly the same than the official source (UIS).
 - Explanations: different definitions of pre-primary education, impossibility of distinguishing between one year and other durations of pre-primary education, difference between pre-primary education and other learning methods...

Table 2. Proportion of pupils who attended pre-primary education

Assessment	Grade	Number of countries (% Developing countries)	Mean	Standard Deviation	Minimum	Maximum
TIMSS 2011	4	35 (40)	74.20	21.86	23.27	99.12
PIRLS 2011	4	51 (41)	76.00	20.89	23.43	99.15
SACMEQ III	6	15 (100)	59.80	25.24	25.65	97.96
TERCE	3	15 (100)	58.25	11.76	37.48	75.91
TERCE	6	15 (100)	58.01	12.66	35.67	77.65
PASEC	2	10 (100)	27.12	11.89	10.90	49.90
PASEC	6	10 (100)	28.21	11.62	12.00	46.60
Mean*		87 (66)	65.48	25.10	11.45	99.15

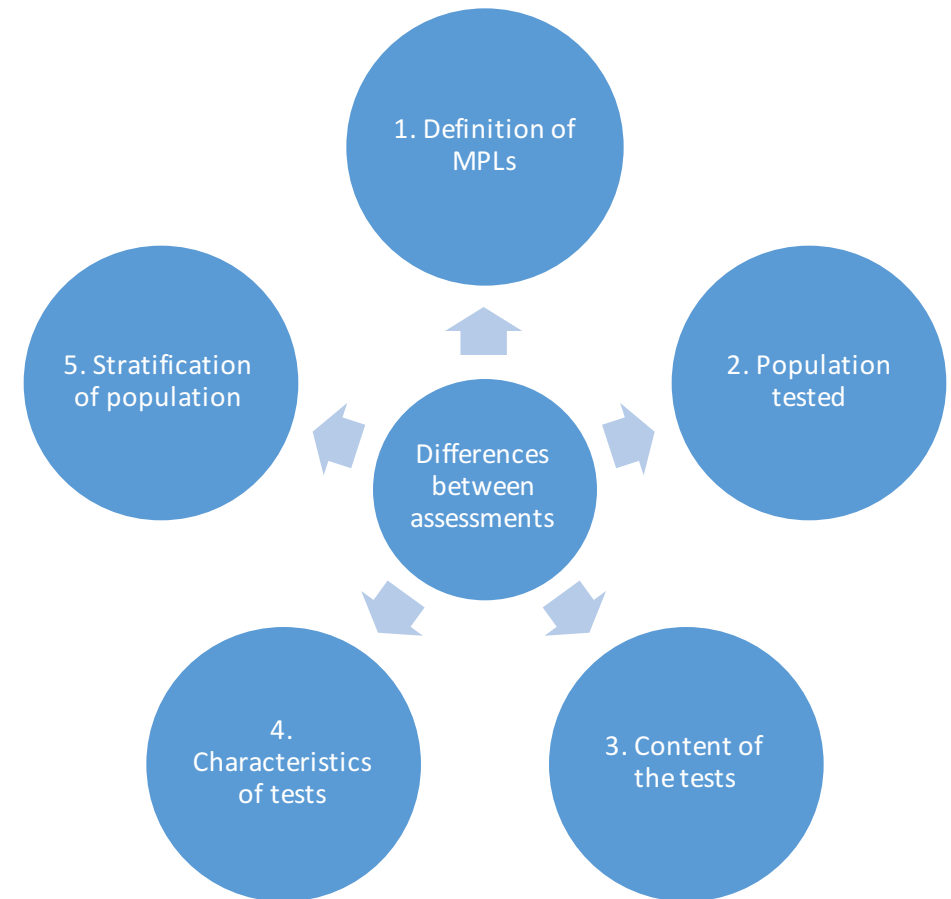
* Since some countries took part to different assessments and grades, a specific computation was made for obtaining the global mean. Mean values are obtained by using the arithmetic mean within each country firstly, regardless to the assessment, and then by computing a global mean of country means.

Figure 8. Comparison between gross enrolment and participation rate to pre-primary education ($R^2 = 0.69$)



Differences between assessments

- Comparison between countries which took part at ISATs and RSATs is not yet possible.
- Five main differences should be highlighted in order to have a better measure in the future
- These differences should be taken into account when a comparison is made between countries which took part at different assessments



Difference 1. Definition of MPLs

- In all assessments, specific proficiency levels are provided.
- The number of proficiency levels differs greatly between assessments (from 3 in PASEC to 8 in SACMEQ).
- The threshold of MPL is not always available and clear in some assessments (example for SACMEQ and TERCE).
- The definition of this threshold differ greatly between assessments, which may conduct to misinterpretations when a comparison is made between assessments.
- Possibility to compare results for countries which took part to at least two different assessments.

Difference 2. Definition of population tested

- With the exception of PISA, all the remaining assessments tests students enrolled in a specific grade.
- In PISA, regardless to the grade, only 15 years old students are tested.
- Thus, depending on drop out ratios, repetition rates, a direct comparison between TIMSS and PISA may conduct to differences for some countries.
- For instance, if in a given country, repetition rates are high in primary and lower secondary education, the students tested in PISA may include classes from grades 6 or 7, while these students are not included in TIMSS assessments.

Difference 3. Content of tests

- The choose of items included inside tests is not a simple process.
- While most assessments focus on curriculums (TIMSS, PIRLS, TERCE and SACMEQ), others give more importance to competences (PISA and PASEC).
- Even if the focus is similar, competences evaluated may differ between assessments: should we consider that competencies required in the life in Sub-Saharan Africa would be the same compared to Developed countries?
- Similarly to competences, assessments which include tests based on curriculums may not be able to be comparable if the coverage of curriculums differ greatly between assessments.

Difference 4. Characteristics of tests

- Tests are not administered simultaneously (different year but also month), with the same duration and the proportion of open-ended questions.
- Most tests are administered between April and June, but specific adjustments are made for the Southern hemisphere countries in assessments like TIMSS.
- The duration of tests may also differ, including the proportion of open-ended questions.
- Some countries are not well prepared for multiple choice questions, which may bias estimation results.
- An important issue relies with the possibility that these differences should impact the results of developing countries where school drop is high and may vary during a school year.

Difference 5. Stratification of population

- Population tested may differ regarding to the definition of the stratification variables.
- While in PISA, schools are primary sampling units, in other assessments, classes are preferred (for example in TIMSS).
- The stratification of the population differs greatly between assessments and may conduct to differences when results for subpopulations are obtained.
- The distinction between urban and rural areas, between public and private schools or between specific regions within countries is possible in some assessments, but not in others (like PASEC or PISA).

Differences highlighted among assessments

- Proportions of children reaching the MPLs differ greatly between assessments (Table 3).
- The lowest proportion is found in PASEC 2014 (both skills), while the highest is observed in PIRLS 2011 (where the proportion of developing countries is the lowest).
- An OLS regression (Table 4) made for confirming these differences show that
 - on average, the PASEC assessment provides proportions of children reaching the MPLs about 30% lower than TIMSS,
 - while the difference is lower but significant for PISA (17%),
 - even if we restrict for countries which took part to at least two different assessments and control for country fixed effects

Table 3. Proportion of Students who reached the minimum proficiency benchmark over assessments

Assessment	Grade	Number of countries (% Developing countries)	Mean	Standard Deviation	Minimum	Maximum
TIMSS 2011	4	55 (44)	80.49	21.96	6.66	99.64
TIMSS 2011	8	49 (61)	73.68	22.25	21.21	98.90
PIRLS 2011	4	54 (39)	85.97	17.18	21.01	99.55
SACMEQ III	6	15 (100)	75.37	16.69	33.29	98.69
TERCE	3	15 (100)	56.66	17.64	15.22	90.00
TERCE	6	15(100)	64.92	19.00	19.94	95.50
PASEC	2	10 (100)	40.74	23.59	9.60	96.70
PASEC	6	10 (100)	41.79	19.92	7.70	86.70
PISA	8	67 (39)	72.57	17.91	24.31	97.26
Mean*		125 (62)	70.96	20.64	8.03	97.90

Notes: Details results for each combination of assessment, grade and skill are available on Table A.7.

* Since some countries took part to different assessments and grades, a specific computation was made for obtaining the global mean. Mean values are obtained by using the arithmetic mean within each country firstly, regardless to the assessment, and then by computing a global mean of country means.

Table 4. Estimation of the differences of proficiency levels between assessments

	(1)	(2)	(3)	(4)	(5)
PIRLS	9.656 (1.807)***	2.219 (2.183)	2.278 (2.340)	6.161 (1.547)***	-3.103 (1.485)**
SACMEQ	-0.942 (4.539)	-4.661 (4.601)	2.117 (6.003)	29.039 (0.891)***	24.407 (0.921)***
TERCE	-14.291 (4.794)***	-18.010 (4.879)***	-12.483 (5.859)**	14.786 (2.108)***	10.154 (2.271)***
PASEC	-35.048 (6.132)***	-38.767 (6.213)***	-30.676 (7.775)***		
PISA	-4.856 (2.806)*	-8.574 (2.939)***	-8.345 (2.899)***	-12.332 (1.212)***	-16.964 (1.326)***
GDP per capita	No	No	Yes	No	No
Dummies for skills	No	Yes	Yes	Yes	Yes
Country dummies	No	No	No	Yes	Yes
Observations	422	422	378	286	286
Countries	125	125	108	68	68
R squared	0.257	0.274	0.303	0.878	0.905

Note: Cluster-robust standard errors in brackets. Clusters are countries. Dummies for TIMSS survey and mathematics used as controls.

Differences highlighted among assessments

- One quick possibility to test for potential differences between assessments can be obtained by comparing the differences of proportions of children reaching the MPLs for countries which took part to at least 2 different assessments.
- Among these countries, most of them took part simultaneously at TIMSS 2011 and PISA 2012.
- Other comparisons can be obtained between SACMEQ III and PIRLS, and between TERCE and PIRLS.
- For instance, about 33% of difference is found in Botswana which took part to both PIRLS 2011 and SACMEQ III
- In other countries, the difference is very small (Honduras).
- No clear relationship between 'doubloon countries' for scaling without any bias assessments

Table 5. Difference of proportion of pupils reaching the MPLs between assessments for countries which took part at two different assessments

Country	Level	Skill	TIMSS	PIRLS	SACMEQ	TERCE	PISA	Variation
Botswana	2	Reading		55.70	89.30			33.60
Kazakhstan	3	Maths	85.41				54.76	30.65
Tunisia	3	Maths	61.47				32.25	29.22
Qatar	3	Maths	53.74				30.44	23.29
Jordan	3	Maths	54.59				31.44	23.16
United Arab Emirates	3	Maths	75.14				53.72	21.42
Israel	3	Maths	86.64				66.50	20.14
Indonesia	3	Maths	43.37				24.31	19.07
Russian Federation	3	Maths	94.86				76.05	18.82
Colombia	2	Reading		72.33		90.30		17.98
USA	3	Maths	91.91				74.15	17.76

Table 5. Difference of proportion of pupils reaching the MPLs between assessments for countries which took part at two different assessments

Country	Level	Skill	TIMSS	PIRLS	SACMEQ	TERCE	PISA	Variation
Japan	3	Maths	96.98				88.94	8.05
Korea Republic of	3	Maths	98.59				90.87	7.73
Singapore	3	Maths	98.90				91.75	7.16
New Zealand	3	Maths	84.13				77.36	6.77
Chile	2	Maths	77.47			83.78		6.32
Hong Kong, China	3	Maths	97.11				91.48	5.62
Honduras	2	Reading		74.26		74.83		0.57
Mean			78.99	64.75	79.50	71.70	66.26	14.51

Differences highlighted among assessments: equity issues

- Theoretically, no specific reason for differences between girls and boys among different assessments (Figure 8.1).
- **Large gaps exist for countries like Indonesia**
 - While girls outperform boys in PISA in Indonesia, the opposite is found for TIMSS.
 - Possible explanations:
 1. Girls may repeat differently than boys and thus populations tested may differ between PISA and TIMSS.
 2. Drop-out may differ between girls and boys at specific grades
 3. Unmeasured education policy may have been conducted in order to impact the gender parity index between 2011 and 2012 (TIMSS = 2011 / PISA = 2012)
- Other differences, often larger ones can be found for Location Parity Index and Wealth Parity Index (Figures 8.2 & 8.3)

Figure 8.1. Comparison of Gender Parity Index between countries which took part at least to two different student achievement tests

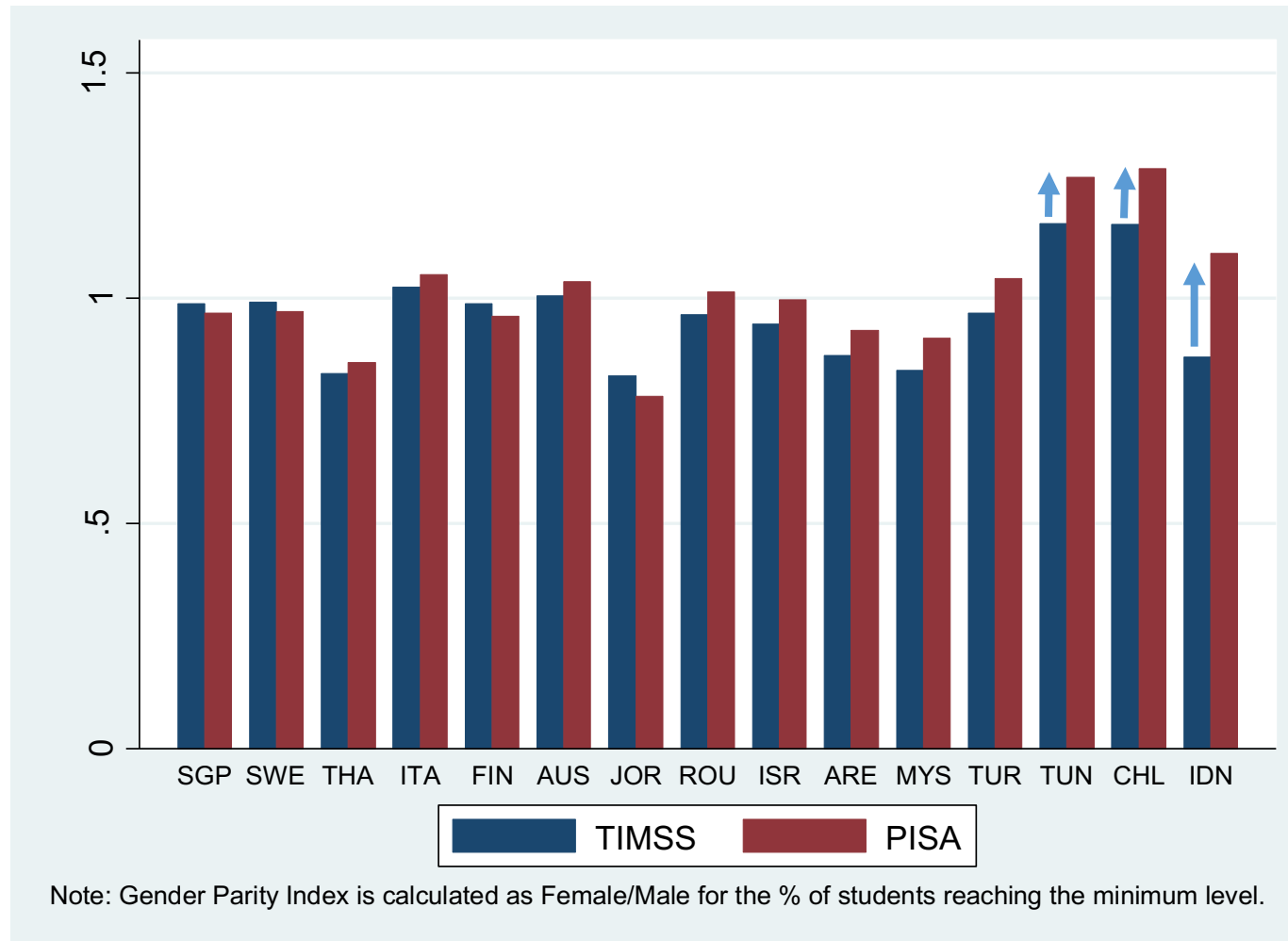


Figure 8.2. Comparison of Location Parity Index between countries which took part at least to two different student achievement tests

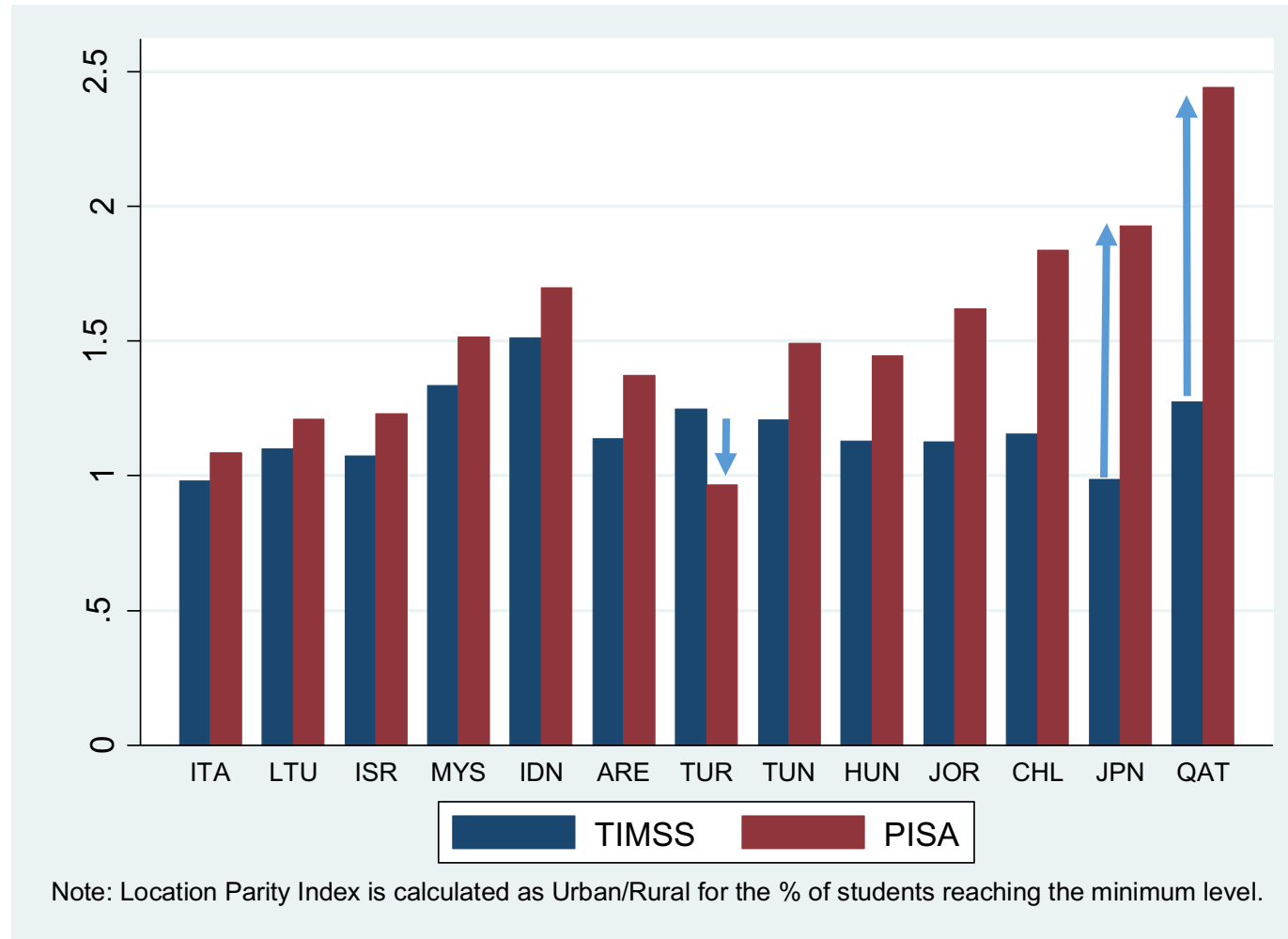
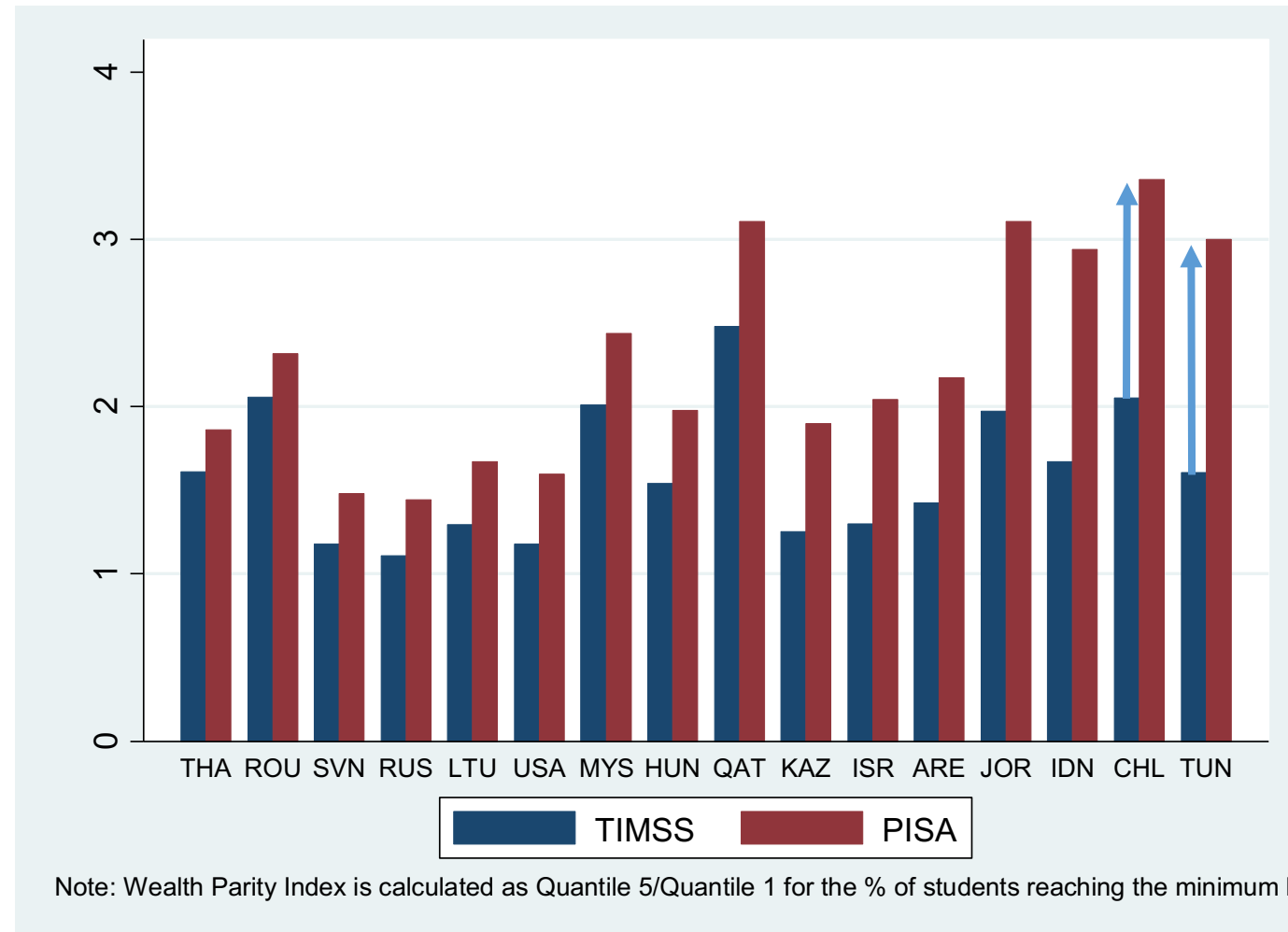


Figure 8.3. Comparison of Wealth Parity Index between countries which took part at least to two different student achievement tests



Short-term objective: a quick measure of MPLs

- A real global metric for learning outcomes will not be available before at least 2022.
- Recent achievement tests must be used to track over time and cross time variations for performance in education.
- All international and regional student achievement tests provide useful information about learning outcomes and more especially about minimum proficiency levels.

Recommendations for future improvements

Step 1

Compare the most recent assessments in order to obtain a first database

Step 2

Methodology for tracking over time within each assessments and extend the database

Step 3

Methodology for the comparison of results between assessments

Step 4

Single indicator for monitoring the SDGs (both quantity and quality of education)

Step 1:

Compare the most recent assessments in order to obtain a global database

- Work done for the UIS in 2016.
- Aggregate all results from TIMSS 2011, PIRLS 2011, PISA 2012, SACMEQ III, TERCE 2013 and PASEC 2014.
- Definition of thresholds for each assessment
- Definition of each inequality groups (wealth, location, gender, language, ethnicity, immigrant status...)
- Include additional measures for other indicators of SDGs (pre-primary education, teacher training, provision of electricity in schools...)

Step 2:

Methodology for tracking over time and extend the database

- Extend the initial database to all existing assessments.
- Analyze the contextual questionnaires for avoiding any change between each round.
- Provide adjustments when needed between each round.
- Data available from 1995 to 2015 (in blue: results are going to be published soon)
 - TIMSS: 1995, 1999, 2003, 2007, 2011 & 2015
 - PIRLS: 2001, 2006, 2011 & 2015
 - PISA: 2000, 2003, 2006, 2009, 2012 & 2015
 - TERCE: 2006, 2013
 - SACMEQ: 1995, 2000, 2007 & 2013
 - PASEC: 1995-2014

Step 3:

Methodology for the comparison of results between assessments

- Existing methodologies used to anchor the assessments
 1. Hanushek & Woessmann (2012): mainly based on NAEP anchoring, but restricted to the ISATs only (very low number of developing countries)
 2. Altinok, de Meulemeester & Diebolt (2014) and Angrist & Patrinos (2013): based on NAEP anchoring but also on double- and low-income countries for RSATs. Provision of data for more than 100 countries.
- Updates which should be done for the SDGs:
 1. Need to update with new released assessments
 2. New data for the MPLs instead of general average scores
 3. Decomposing data for several subpopulations (gender, location...).
 4. Make simulations for learning outcomes until 2030 (see for instance Education Commission's report on learning outcomes)

Step 4:

Methodology for the comparison of results between assessments

- EFA Dakar in 2000 : « Education for All »
- SDGs Incheon in 2015: « Quality Education for All »: need to focus on both access, completion and high quality education
- Proposal of the creation of an hybrid indicator which may focus on the proportion of children who achieve a given level of education with the minimum proficiency skills in both mathematics and reading.
 - Propose this indicator for both primary and lower secondary education.
 - Add an equity dimension based on gender/location/language/wealth topics.

Main bibliography

- Upcoming report & database:
 - ALTINOK for the UIS: « The contribution of student achievement tests to the monitoring of sustainable development goals »
- General documents:
 - UIS/UNESCO (2016), « Sustainable Development Digest », [link](#)
 - Education Commission (2016), « The Learning Generation », [link](#)
- Existing datasets for tracking learning outcomes
 - Hanushek & Woessmann (2012), « Do better schools lead to more growth? », [link](#)
 - Altinok, Diebolt & de Meulemeester (2014), « A new database on education quality », [link](#)
 - Angrist, Patrinos, Schlotter (2013), « An expansion of a global data set on educational quality », [link](#)