JULY 2025

Data

National primary-level learning proficiency statistics from Africa

A stocktaking and hybrid harmonisation

Author:
Martin Gustafsson (RESEP, University of Stellenbosch)

AFLEARN DATAHUB

National primary-level learning proficiency statistics from Africa

A stocktaking and hybrid harmonisation

Martin Gustafsson
31 August 2025

ABSTRACT

This paper responds to three needs. First is a need for updated and more comprehensive and comparable country-level proficiency statistics for the primary schooling level in Africa. Secondly, a systematic and updated review of the relevant data and statistics available for Africa, and their levels of reliability, seemed necessary. Third is the need to explore new ways of harmonising learning proficiency statistics across countries. With regard to the first and second needs, nine data sources are critically discussed, each covering between 2 and 18 countries on the African continent. Seven of the nine programmes involve testing in schools and two in households. By comparing microdata across programmes, where microdata were available, but also comparing national statistics in the absence of microdata, harmonised statistics were produced for 47 of 55 African countries, representing 97% of the continent's children. The analysis, which attempts to balance coverage with rigour, confirms that both data quality problems and methodological issues make harmonisation difficult. The more coverage achieved, the greater the reliability concerns. Arguably the value of the paper lies at least as much in the evaluation of the data from the programmes, programmes which are continually evolving, as from the final harmonised statistics. Regarding the need to explore new methods, the paper pays special attention to one issue which has probably received too little attention in past harmonisations: adjustments to take into account the fact that even within the same programme different grades might be tested. Moreover, the paper suggests that even published statistics where the underlying microdata are not available for analysis are worth considering in a harmonisation exercise.

Contents

Tables and figures

Country ISO codes

AGO	Angola
BDI	Burundi
BEN	Benin
BFA	Burkina Faso
BWA	Botswana
CAF	Central African Republic
CIV	Côte d'Ivoire
CMR	Cameroon
COD	DRC (Democratic Republic of the Congo)
COG	Congo
COM	Comoros
CPV	Cape Verde
DJI	Djibouti
DZA	Algeria
EGY	Egypt
ERI	Eritrea
ESH	Sahrawi Republic
ETH	Ethiopia
GAB	Gabon
GHA	Ghana
GIN	Guinea
GMB	Gambia
GNB	Guinea-Bissau
GNQ	Equatorial Guinea
KEN	Kenya
LBR	Liberia
LBY	Libya
LSO	Lesotho
MAR	Morocco
MDG	Madagascar
MLI	Mali
MOZ	Mozambique
MRT	Mauritania
MUS	Mauritius
MWI	Malawi
NAM	Namibia
NER	Niger
NGA	Nigeria
RWA	Rwanda
SDN	Sudan
SEN	Senegal
SLE	Sierra Leone
SOM	Somalia
SSD	South Sudan
STP	São Tomé and Príncipe
SWZ	Eswatini
SYC	Seychelles
TCD	Chad
TGO	Togo
TUN	Tunisia
TZA	Tanzania
UGA	Uganda
ZAF	South Africa
ZMB	Zambia
ZWE	Zimbabwe

EXECUTIVE SUMMARY

Section 1 introduces the paper, whose aim is in part to take is to take stock of what data and statistics exist on learning outcomes at the primary level. Data currently cover countries representing 79% to 97% of Africa's children, with the exact percentage depending on what criteria for data quality are used. The paper moreover produces a new set of harmonised end of primary proficiency statistics for African countries, for reading and mathematics, using definitions of proficiency put forward by UNESCO for the purposes of Sustainable Development Goals (SDG) reporting.

Section 2 presents the stocktaking, with section 2.1 identifying nine assessment programmes which between them generate a set of 145 national statistics spanning 47 countries.

Section 2.2 describes the nine assessment programmes, including indicators of their reliability and their alignment to SDG standards. Roughly in order from greater to lesser ability to produce the desired proficiency statistics, the programmes are: TIMSS; PIRLS; LaNA; AMPL; PASEC; SACMEQ; MICS; PAL; and EGRA. TIMSS and PIRLS are longstanding programmes frequently used for SGB reporting. Their limited footprint in Africa has in part been addressed through the sister programme LaNA, which began collecting learning data in 2023. AMPL was initiated by UNESCO in 2021 and specifically aims to produce the required proficiency values. PASEC and SACMEQ (recently rebranded as SEACMEQ) are large African programmes focussing largely on francophone and anglophone countries respectively. Published PASEC values are more recent than those for SACMEQ, and are considered more reliable, in part because of higher levels of transparency with respect to technical documentation and the availability of microdata. MICS is a longstanding UNICEF programme collecting data from households which since 2017 has included an assessment module for children. MICS expands substantially the number of countries with learning data. PAL is another household-focussed programme with a learning assessment, focussing on a small group of countries. Finally, EGRA involves assessing children in schools on reading. Its limitations for the current study include that its focus on across-country comparability is not strong, and that its microdata are not readily available. Moreover, the future of EGRA appears uncertain following USAID's closure, and given that EGRA has relied strongly on USAID funding.

Section 2.3 represents the data availability situation graphically. Section 2.4 discusses the confidence intervals around proficiency statistics which are likely to apply in the final harmonised sets of values. In the interests of simplicity, what the current work avoids is discussing implications for confidence intervals at each step of the harmonisation process, something which would be complex given the hybrid nature of the harmonisation.

Section 3 describes the harmonisation, with section 3.1 covering the extent to which there are linking countries which are common across the nine programmes. The presence of such countries appears good. The sub-section also discusses key principles of the hybrid harmonisation process. Above all, it is recognised that the process is intended to respond to the needs of education specialists, who are likely to have an interest in the situation of individual countries, both with respect to estimated proficiency levels and the quality of the available data. Past harmonisations have tended to be more oriented towards facilitating averages across countries and producing data that could be used in economic analysis. In line with the need, the approach to be followed does not include imputing proficiency levels on the basis of non-assessment data, for instance income or enrolment data. Such imputation has been pursued in past harmonisation work. Moreover, a requirement used in past work is dropped, namely that learning statistics in the absence of the underlying microdata should not be used. The avoidance of imputations based on non-assessment data would narrow the set of countries for the current work, while the embracing of learning statistics without microdata would widen the set of countries.

Section 3.2 discusses existing and new evidence on grade-on-grade learning gains among learners in Africa with a view to informing grade-related adjustments. Patterns are relatively consistent, which facilitates the implementation of adjustments to proficiency statistics where within a programme a non-standard grade was tested. For example, in AMPL Zambia tested grade 7, where other countries tested grade 6. This necessitates a downward adjustment of Zambia's proficiency value, in the interests of comparability. Section 3.3 presents the results of implementing grade-related adjustments for eight countries in the programmes LaNA, AMPL, PASEC and TIMSS.

Section 3.4 explores the existence of clearly implausible values, essentially instances where the rankings of individual countries are excessively different in different programmes. Instances necessitating a removal of values were not found.

Section 3.5 aligns values among four of the nine programmes considered to conform relatively well with UNESCO standards, namely PIRLS, TIMSS, LaNA and AMPL. In the case of four countries spanning more than one programme, simple across-programme averages within each of the two subjects are calculated. Across-subject averages per country are also calculated. This helps to smoothen out the anomaly resulting from different programmes balancing difficulty levels differently across the two subjects. The anomaly means that a country can appear to perform far better in mathematics than reading in one programme, but with the opposite picture emerging in another programme. The result of this step in the harmonisation process is 13 countries, each with harmonised proficiency values for mathematics and reading, and for the average across the two.

Section 3.6 deals with the complex task of harmonising statistics from the remaining five programmes considered to reflect UNESCO standards less reliably. There are four linking countries common across SACMEQ and the abovementioned 13-country set considered the standard. The original SACMEQ-derived proficiency statistics, covering 13 countries, are clearly derived more leniently than those in the standard, and are thus adjusted downwards using a linear regression. No microdata for SACMEQ were available to explore distributions of skills in more depth.

In the case of 14 PASEC countries, of which four were also found in the standard, microdata were available. This allowed a traditional microdata-based harmonisation of PASEC to the standard in the case of reading for two countries which also had LaNA microdata, namely Senegal and Burkina Faso. In the case of mathematics, however, the microdata approach produced proficiency statistics for the 14 PASEC countries which were clearly too low, suggesting the two linking countries were atypical. Thus, for mathematics a non-microdata approach of simply seeking the optimal alignment of national statistics was pursued.

MICS microdata from 15 countries were analysed. These data, collected from households, are valuable insofar as they permit insights into the extent to which assessing learners in the end of primary at any point in time might under-estimate learners who eventually reach the relevant proficiency levels at school. This matter is important in Africa, where grade repetition levels are often very high. In terms of the harmonisation process, the simple approach is taken of considering assessment results for children enrolled in the last primary grade. Only the microdata for the numeracy assessments are used, given relatively high levels of non-participation in the reading assessments. For estimates of reading, predictions based on the numeracy data are used. A further issue is that MICS assessments are designed primarily with early grade learners in mind, meaning that there are not enough difficult questions to allow for a meaningful harmonisation based on a comparison of MICS and, say, PASEC data. MICS estimates are therefore aligned to the standard without using the microdata and only by comparing country aggregates.

Published PAL statistics, which are also based on household data, are used to derive harmonised values for just one country, namely Tanzania.

Published EGRA mean values are used to produce percentage proficient values for seven countries with no other data source. In doing this, the relationship between the seven countries and the rest of the continent as seen in earlier Angrist *et al* (2021) is replicated in the new set of harmonised values. Fortuitously, this process resolves certain country ranking anomalies with respect to the seven countries in the new set of statistics.

Section 3.7 concludes the harmonisation by selecting an optimum set of statistics per country where more than one exists. These statistics, which cover countries representing 97% of Africa's children, indicate that only 13.4% of the continent's learners reach SDG-defined proficiency levels at the end of primary. This is considering the average across reading and mathematics. Specialists who viewed the new set of harmonised statistics found them to be plausible on the whole, but raised questions in relation to the ranking of certain countries. Reasons why these countries are ranked as they are, and reasons for considering these rankings plausible, are presented.

Section 4 finds that the continent's 13.4% proficiency level for enrolled learners drops to 10.8% if children not in school are taken into account, using the common assumption that the out-of-school do not reach desired proficiency levels.

Section 5 uses MICS data to examine the extent to which learning levels at the end of primary are predicted by levels of learning in the earliest school grades. It is confirmed that learning already in grade 1 is highly predictive of learning in later grades. Strikingly, the gains made across grades within countries are considerably smaller than the stark differences in proficiency levels that exist across countries already in grade 1. A question is what lies behind the grade 1 differences. A comprehensive answer is outside the scope of the paper, yet MICS data suggest that much of the explanation lies in a 'burst' of skills obtained already in the first grade, with some countries being much better at this than others.

Section 6 discusses the importance of monitoring trends over time per country, as opposed to just comparing countries to each other. Both are important, but the SDGs are ultimately concerned with improving education over time. Reporting on improvements in Africa, including improvements against SDG education indicators is too infrequent and when it occurs tends to pay more attention to the large international programmes with high levels of transparency in terms of technical reports and the availability of the microdata to researchers. The status and even sustainability of African programmes such as PASEC and SACMEQ, whose role in monitoring trends is vital, could be improved through better and more readily available technical products.

Section 7 concludes the paper. It is confirmed that with the new analysis presented in the paper half of Africa's countries were given harmonised measures of proficiency using data which had not been used before for harmonisation purposes. The paper thus represents a substantial updating of indicator values, something made possible in large part due to three new data collections: MICS foundational learning; LaNA; and AMPL. Pointers for future harmonisation work are provided, as well as suggestions on improving the availability of microdata and tightening definitions of proficiency.

1 Introduction

Despite some evidence of good educational practices in Africa and countries displaying improvements, on average Africa displays very low levels of learning proficiency[1]. The commitment to addressing this problem, and learning from cases where success has been achieved, can be seen in for instance the African Union's Nouakchott Declaration, which states that Africa should 'end learning poverty by 2035' in the early school grades. While this is an extremely ambitious aim, given how gradually educational improvement has occurred in even successful countries, improving learning outcomes substantially by 2035 is possible.

The current paper presents a picture of what we know about learning proficiency at the primary level across Africa. Many important monitoring programmes have developed in the last couple of decades. The continent has reached a critical point where between 79% and 97% of school-age children are in countries where there is at least some measure of learning from an international programme allowing for comparison across countries. The exact percentage depends on the definition of what can be considered sufficiently reliable data. The current level of coverage is a major achievement. However, for various historical reasons, often related to language and regional groupings on the continent, it has been difficult to compare statistics across different monitoring programmes.

Thus, in addition to describing the statistics we have the paper aligns, or harmonises, statistics across programmes but also within programmes where this is deemed necessary. The aim is to produce a clearer picture of how successful different African countries are at achieving effective learning at the primary level. This, in turn, helps to bring about a more informed discussion of how to deal with the challenges.

Specifically, the paper aims to produce best possible estimates of the percentage of learners at the end of primary who have reached an adequate level of learning proficiency in reading and mathematics. This is done for 47 of 55 African countries, where these 47 countries cover 97% of Africa's children aged 5 to 14. Adequacy with respect to learning outcomes is understood in terms of the Sustainable Development Goal (SDG) indicator 4.1.1b. Despite this focus on the end of primary, statistics and data from all grades within the primary level are considered. This means, for instance, that information on learning in the earlier grades is used to inform, or predict, levels of learning at the end of primary. Producing harmonised values for the end of primary is in several respects more feasible than doing this for the initial grades. Not only have data for the end of primary displayed important advances in recent years within the continent, what to consider a minimum standard is clearer at this level, though as will be discussed in this paper, even here certain issues could be clearer. While important advances are also being made with respect to the rather different process of generating comparable measures of learning for the initial three grades, several methodological issues at this level are still in the process of being resolved[2].

Of the 145 national statistics constituting the backbone of the paper's analysis, 91% are from 2013 or later. And 94% of the 47 countries with statistics have a statistic that is not older than 2013.

The analysis and discussion necessitated by the harmonisation should in itself be of use as it brings to the fore assessment design and calculation issues which warrant attention among education researchers and planners.

Section 2 describes, in part through a series of visualisations, existing statistics on learning outcomes. Details for each of nine monitoring programmes whose statistics seemed usable are

[1] UNESCO Institute for Statistics, 2024a.
[2] See for instance UNESCO Institute for Statistics (2024c).

discussed. Critically, this includes some reflection on the reliability of each programme's statistics, which is in part related to the transparency of each programme in terms of the availability and use of the programme's microdata, and the depth and maturity of each programme's technical reports.

Section 3 presents a harmonisation. As a preliminary step, the question of how much learning is typically gained by learners in a year is assessed, using the existing literature and analysis of MICS data. This information is needed for adjustments where available statistics refer to different grades. Thereafter, the harmonisation follows five steps. Step 1 involves adjustments within certain assessment programmes, in particular where grades are not aligned. What is not done is to adjust values using assumptions on progress in the schooling system between the year of the data collection and today. The variation across countries in terms of progress is likely to be large, and there is little data available to gauge change over time. Step 2 involves examining whether certain statistics stand out as being clearly implausible and should be excluded. Step 3 involves identifying programmes, among the nine, whose values can be considered to reflect the end of primary minimum proficiency standard with a relatively high degree of accuracy. These values become the 'anchors' for the next step. In step 4, proficiency values which are not in line with the standard are adjusted so that alignment is achieved. The adjustment is effected by using linking countries participating in more than one programme and then adjusting the values of linking and non-linking countries within non-standard programmes to align with values in the standard. In this step microdata is used when available, though even without this data adjustments may be attempted. Step 5 involves selecting the most plausible statistic per country, where more than one is available, and assessing the credibility of selected statistics through comparison to earlier harmonisation work.

The term 'hybrid harmonisation' has been used in this paper, for instance in its title, to distinguish what is undertaken here from earlier harmonisation work. That harmonisation work carries the advantage of relying strongly on the microdata from programmes, which improves the precision of harmonised statistics. Moreover, past work has all been global, not region-specific, meaning the earlier harmonisation work has drawn from the widest possible range of overlaps between programmes. What is presented here is region-specific, and involves a mix of approaches, with microdata being used only partially. What appears not to have been done previously is to implement upward or downward adjustment where, within a programme, not all countries tested the same grade. Arguably, one of the advantages of the approach taken in this paper is that it widens the pool of countries with values by including proficiency statistics even if the microdata are not available, where earlier harmonisations have tended to widen the pool by estimating learning on the basis of *non*-assessment data, such as income per capita.

Section 4 reflects on the extent to which children do not reach the last grade of primary school across Africa, and how this biases the picture emerging if only enrolled children are considered.

Section 5 expands the MICS analysis of grade-on-grade learning gains by visualising how dependent educational success is on what occurs in the very earliest grades.

The central focus of the paper is on comparing countries to each other in better ways. Yet ultimately comparing trends in one country over time is even more important for effective education planning. Section 6 discusses challenges in obtaining credible trends over time, but also instances where gains in proficiency in African countries have been globally acknowledged.

Finally, section 7 concludes, while reflecting on how the kind of analysis undertaken for this paper can be taken forward. It also discusses pointers for future data collection.

2 A stocktaking of existing statistics for Africa

2.1 The available statistics

Table 1 below was compiled following some exploration of which international, or cross-country, assessment programmes appeared relevant for an Africa-focussed attempt to arrive at best possible estimates of learning proficiency at the end of primary. Nine programmes are reflected in the table. The nine acronyms and associated full names are as follows:

AMPL	Assessment for Minimum Proficiency Level
EGRA	Early Grade Reading Assessment
LaNA	Literacy and Numeracy Assessment
MICS	Multiple Indicator Cluster Surveys
PAL	People's Action for Learning
PASEC	Programme d'analyse des systèmes éducatifs de la Confemen (English: Program for the Analysis of Educational Systems of CONFEMEN)
PIRLS	Progress in International Reading Literacy Study
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality[3]
TIMSS	Trends in International Mathematics and Science Study

What is not within the scope of the current paper is data collected at the secondary level, such as PISA[4] for Development (PISA-D). In Africa such data are less available than for the primary level and would not have covered countries which were not already covered by the primary level data. That said, future analyses could benefit from the inclusion of data from the secondary level, as these data can help to verify findings at the primary level.

To a limited extent, data from national assessment programmes were examined to see whether they might add value to the analysis. For this to be possible, a national programme should ideally be anchored to some international assessment through shared items. At face value, the most promising national assessment was that of Rwanda implemented in 2023 for grades 3 and 6[5]. However, alignment to UNESCO standards was pursued through a non-statistical approach, and not through common items. This is not necessarily a problem, but the resultant proficiency statistics were so high that it is very unlikely that standards typically used in the international programmes applied. This is discussed further in section 3.6.4.

Table 1 presents learning outcomes statistics from the nine selected programmes which have been published somewhere and which serve as a point of departure for this paper's harmonisation exercise. Statistics for 47 countries were found. EGRA values are different insofar as they are test score means, while all other values are proficiency statistics indicating the extent to which children reach some competency threshold. If the acronym is followed by 'r', the values refer to reading competencies, while 'm' indicates mathematics competencies.

[3] The official acronym is currently SEACMEQ, but the applicable acronym in 2013, when the data informing the current paper were collected, was still SACMEQ.
[4] Programme for International Student Assessment.
[5] Rwanda: National Examination and School Inspection Authority, 2023.

Table 1: Values by country and assessment

	MICSr	MICSn	PASECr	PASECm	SACMEQr	SACMEQm	LaNAr	LaNAm	TIMSS	PALr	PALm	AMPLr	AMPLm	PIRLS	EGRA
AGO															263
BDI			28.2	60.9								0.1	13.5		323
BEN	14.2	17.0	75.0	51.6											
BFA			66.7	62.5			24.0	13.0				9.0	23.7		
BWA					68.5	36.6									
CAF	4.7	1.5													
CIV			40.4	17.2								10.8	8.9		
CMR			53.6	32.9											
COD	8.7	0.5	27.1	18.4											248*
COG			58.4	33.4											
COM	25.6	25.5													
DZA									14.0						
EGY							42.0	29.0						45.0	333
ETH															324*
GAB			93.4	66.7											
GHA	21.4	15.7													229
GIN			44.7	32.4											
GMB	12.4	8.1													285
GNB	12.4	7.5													
KEN					77.6	53.2				92.0	90.0	25.2	36.9		364*
LBR															274*
LSO	43.8	13.3			48.4	9.5						10.8	19.7		
MAR									22.0						268*
MDG	23.3	7.3	17.5	21.6											
MLI															229*
MOZ					36.3	15.2									
MUS					75.3	58.9									
MWI	18.8	12.6			15.3	4.1									275
NAM					61.3	17.5									
NER			30.0	22.5											
NGA	26.8	25.3					20.0	11.0							262*
RWA															321

	MICSr	MICSn	PASECr	PASECm	SACMEQr	SACMEQm	LaNAr	LaNAm	TIMSS	PALr	PALm	AMPLr	AMPLm	PIRLS	EGRA
SDN															361
SEN			74.7	65.0			12.0	8.0				13.3	34.0		267
SLE	16.0	1.6													245
SSD															281*
STP	38.4	36.0													
SWZ	49.2	28.9			84.3	37.5									
SYC					79.7	52.4									
TCD	4.4	6.8	22.1	11.5											
TGO	17.7	8.8	38.9	37.0											
TUN	66.0	28.2							9.0						
TZA										86.0	77.0				377
UGA					50.6	20.7				63.0	70.0				227*
ZAF					57.2	29.7			17.0					19.0	
ZMB					20.8	7.1						9.7	16.0		229
ZWE	44.4	24.6			45.4	23.4									
Count	18	18	14	14	13	13	4	4	4	3	3	7	7	2	21

Table 2 indicates the year in which the data collections behind the Table 1 statistics occurred. In some instances data collection occurred over two consecutive calendar years, possibly because the school year spans two calendar years. In such cases, the second of the two calendar years was placed into Table 2.

Table 2: Data collection year by country and assessment

	MICS	PASEC	SACMEQ	LaNA	TIMSS	PAL	AMPL	PIRLS	EGRA	Last
AGO									2011	2011
BDI		2019					2021		2012	2021
BEN	2022	2019								2022
BFA		2019		2023			2021			2023
BWA			2013							2013
CAF	2019									2019
CIV		2019					2021			2021
CMR		2019								2019
COD	2018	2019							2012	2019
COG		2019								2019
COM	2022									2022
DZA					2007					2007
EGY				2023				2021	2013	2023
ETH									2010	2010
GAB		2019								2019
GHA	2018								2013	2018
GIN		2019								2019
GMB	2018								2011	2018
GNB	2019									2019
KEN			2013			2015	2023		2013	2023
LBR									2013	2013
LSO	2018		2013				2023			2023
MAR					2023				2011	2023
MDG	2018	2019								2019
MLI									2015	2015
MOZ			2013							2013
MUS			2013							2013
MWI	2020		2013						2012	2020
NAM			2013							2013
NER		2019								2019
NGA	2021			2023					2010	2023
RWA									2016	2016
SDN									2015	2015
SEN		2019		2023			2021		2009	2023
SLE	2017								2014	2017
SSD									2017	2017
STP	2019									2019
SWZ	2022		2013							2022
SYC			2013							2013
TCD	2019	2019								2019
TGO	2017	2019								2019
TUN	2018				2007					2018
TZA						2015			2013	2015
UGA			2013			2015			2009	2015
ZAF			2013		2023			2021		2023
ZMB			2013				2023		2011	2023
ZWE	2019		2013							2019

To illustrate the slowness of improvement in learning outcomes, a phenomenon which makes it feasible to use relatively old statistics for some countries. PASEC 2014 and 2019 proficiency statistics can be compared. Across the ten PASEC countries with results in both years, the gain in the percentage of learners who were proficient in mathematics at the end of primary was *minus* 0.5 percentage points a year[6].

Of the 145 national statistics appearing in Table 1, 91% are from 2013 or later. And 94% of the 47 countries with statistics have a statistic that is not older than 2013.

2.2 Characteristics of the nine programmes

Below, the characteristics of the nine programmes are discussed, particularly with respect to issues relevant for the harmonisation work. Briefly, how are the underlying data collected? When were they collected, and would they have been affected by the COVID-19 pandemic? If schools, as opposed to households, were surveyed, what grade was tested[7]? Where were the values from Table 1 taken from? What language was used in the testing? Is good metadata documentation publicly available? In the case of the proficiency statistics, how was the threshold determined? To what extent is the threshold in line with existing standards for end of primary set by UNESCO? Are the microdata easily available? And what are the reliability or quality concerns?

The sequence of the programmes described below is very roughly from more to less able to provide standardised statistics aligned to the UNESCO standards. The descriptions confirm the desirability of aiming to estimate harmonised statistics for the end of primary, as opposed to the initial school grades.

TIMSS is a longstanding programme of the IEA[8] with excellent documentation and microdata that are easily accessed. TIMSS has been used to define a minimum level of proficiency in mathematics for SDG indicator 4.1.1b, which deals with the end of primary. A 2021 UNESCO guide[9] stipulates that what TIMSS defines as the 'intermediate benchmark' should be considered the minimum standard for mathematics at the end of primary. The four TIMSS values in Table 1 are the most recent proficiency statistics, using the intermediate benchmark, in the official TIMSS international reports. This benchmark is 475 TIMSS points. For Algeria and Tunisia values are from 2007 testing, while for South Africa and Morocco far more recent 2023 values were available. For the latter two countries the pandemic is likely to have exerted an influence on the values. All countries assessed grade 4, except South Africa, which assessed grade 5.

There is obviously a potentially large risk inherent in using statistics which are as old as the 2007 ones. If, say, Tunisia improved continuously, over the 2007 to 2025 period, as fast as Morocco did between 2015 and 2023, Tunisia's 9% of 2007 would become around 45% in 2025. In such a case, the 9% would clearly be deceptive. However, improvements such as Morocco's are highly unusual, and global proficiency data from 2000 onwards suggest that on average the world has seen virtually no improvement[10]. Morocco, and South Africa, have been found to be unusual for the steepness of their improvements[11]. Nonetheless, the age of the 2007

[6] 2014 values from CONFEMEN (2014). Annual improvements over the five year period ranged from 3.0 percentage points for Niger to minus 5.2 for Burundi. It should be remembered that trends over time should be interpreted with even greater care than trends across countries within one year, given dynamics such as changing school participation levels and changes in the time of year when learners were tested.

[7] One complexity not dealt with here is the fact that testing occurring at the start of a school year would yield different results to testing at the end of the school year. Instead, only the grade tested is considered.

[8] International Association for the Evaluation of Educational Achievement.

[9] UNESCO Institute for Statistics, 2021a.

[10] UNESCO Institute for Statistics, 2024a.

[11] McKinsey & Company, 2024.

statistics suggests it is important to consider other statistics, which is possible for Tunisia, though Algeria has no alternative in Table 1. South Africa tested using English and Afrikaans, while the other countries used Arabic.

Crucially, the UIS approach is to consider, for example, 22% of end of primary learners being proficient in mathematics *if* 22% *of grade 4 learners* reach at least the intermediate benchmark. Something is assumed regarding a later grade based on the situation in an earlier grade. This is reflected in the fact that the official UNESCO SDG monitoring system gives a 22% value for indicator 4.1.1b for mathematics in the case of Morocco. The last grade of primary in Morocco is grade 6. This approach is obviously different from taking the intermediate benchmark standard, and then assuming learners gain a specific amount of learning up to grade 6, and then estimating a grade 6 value, which would clearly be greater than 22%. The assumed UIS thinking is not made explicit, but it is implied, and this influences the way adjustments are made, or not made, in the sections that follow.

What would require an adjustment is South Africa's TIMSS statistic, which is too high in the sense that grade 5 was tested, as opposed to the standard grade, which is grade 4.

The advantage with TIMSS (and PIRLS) is that the long history of the programme has allowed lessons to be learnt, and that the transparency of the programme instils trust. A disadvantage for the current work is that the footprint of TIMSS is limited within Africa. Moreover, the use of a performance threshold in grade 4 to implicitly predict proficiency in grade 6 can be confusing.

PIRLS is also an IEA programme, similar in design to TIMSS, except its focus is reading. A key difference is that the abovementioned 2021 UNESCO guide stipulates that the *low* (not intermediate) benchmark in PIRLS should be used for indicator 4.1.1b with respect to reading. The grade issue in PIRLS is similar to that in TIMSS. Thus Egypt's 45% reading proficiency value obtained from testing grade 4 learners in 2021 is used for end of primary in grade 6 in the SDG reporting system. The same applies to South Africa. Both PIRLS values in Table 1 are from 2021 PIRLS testing, meaning pandemic effects are very likely, though also quite limited in magnitude. To illustrate, South Africa's value moved from 21% in PIRLS 2016 to the 19% seen in 2021. South Africa tested using all its 11 official languages, while the other countries used Arabic.

The advantages and disadvantages of PIRLS for the current work are essentially the same as for TIMSS.

LaNA is the third IEA programme in Table 1. This programme is a new one, designed specifically for developing countries with low levels of proficiency, though LaNA produces statistics which are on the same scale as TIMSS (for mathematics) and PIRLS (for reading). The release of the first round of LaNA statistics in 2025[12], following testing in 2023, takes our knowledge of learning in Africa forward substantially. Microdata are now available online. The fact that testing occurred in 2023 means pandemic-related learning losses could still have been present. Of the four LaNA countries shown in Table 1, three tested grade 6, while Egypt tested grade 5. The LaNA proficiency statistics of Table 1 are taken from tables in the LaNA report which use the PIRLS low benchmark and TIMSS intermediate benchmark[13]. It can be assumed that in order to realise full comparability to the abovementioned TIMSS and PIRLS statistics, downward adjustments would be necessary so that the standard of using grade 4 were simulated. In other words, an adjustment similar to that applicable to South Africa in TIMSS would need to be applied.

[12] Von Davier *et al*, 2025.
[13] Von Davier *et al*, 2025: 14, 19.

Outside of Africa, LaNA covered Pakistan and Palestine. Languages used for testing in LaNA are not specified in the 2025 LaNA report, but the microdata indicate English in Nigeria, French in Burkina Faso and Senegal, and Arabic in Egypt. The Nigerian sample excluded around 10% of the grade 6 population, primarily learners in Islamic religious schools in the north of the country. In fact, closer examination of the documentation of several programmes reveals some exclusions, though seldom as high as 10%. In particular, very small and remote schools are commonly excluded during sampling. This has been true in for example SACMEQ.

Because LaNA uses a proficiency threshold originally applicable to grade 4 for grades above grade 4, in the case of the four African LaNA countries, and the need for adjustments that arises from this, complicate the use of LaNA for the current analysis. Nonetheless, the abovementioned advantages of transparency and programmatic maturity applicable to TIMSS and PIRLS largely apply to LaNA.

AMPL is a relatively new assessment programme developed jointly by the UNESCO Institute for Statistics (UIS) and the Australian Council for Educational Research (ACER). It has concentrated both on assessment at the lower primary level, known as AMPLa, and the end of primary, known as AMPLb. The programme was initially in large part a response to the need to understand learning losses in Africa during the pandemic. A 2021 wave of testing at around the upper primary level occurred in: Burkina Faso grade 6, Burundi grade 6, Côte d'Ivoire grade 6, Kenya grade 7, Senegal grade 6, and Zambia grade 5[14]. A separate 2023 wave of testing focussed, at the upper primary level, on Lesotho grade 7, Kenya grade 6 and Zambia grade 7[15]. Test booklets were available for English and French only. There were thus seven countries for which values could be inserted in Table 1. Choices between the 2021 and 2023 values had to be made for Kenya and Zambia. For both these countries the 2023 values were chosen as they corresponded to the last grade of primary – grades 6 and 7 respectively. Surprisingly, across the six countries pandemic-related learning losses were barely found following the 2021 collection – they could only be detected for reading in Kenya. AMPL proficiency thresholds were designed in line with the end of primary indicator standards defined by UNESCO and AMPL aimed to reproduce thresholds seen in TIMSS and PIRLS. The AMPL microdata have been made available on the UIS website[16].

One clear advantage for the current analysis with AMPL is that it is the only programme used here which was specifically designed to gauge proficiency in terms of minimum competencies identified by UNESCO for the purposes of monitoring SDG goal 4. In the case of several other programmes, alignment to a UNESCO standard was pursued, but retroactively and largely on the basis of thresholds that existed prior to the SDGs. Though AMPL has focussed on Africa, it has also been used in a non-national sample in India.

PASEC. PASEC has focussed mainly on nationally representative samples of schools across African francophone countries and has reported on lower primary and end of primary trends in reading and mathematics across 2014 and 2019. The values for the 14 PASEC countries in Table 1 are 2019 values seen in the official 2019 report[17]. The grade tested is grade 6 for most countries, though grade 5 in the case of Madagascar and Gabon, both countries where primary schooling ends in that grade[18]. For end of primary, tests in French were used across all countries, though alternatives were also used in two countries: Arabic in Chad and English in Cameroon[19].

[14] UNESCO Institute for Statistics, 2022a: 21, 79.
[15] UNESCO Institute for Statistics, 2024b: 9.
[16] https://ampl.uis.unesco.org/amplab-databases-and-analysis.
[17] CONFEMEN, 2020a.
[18] CONFEMEN (2020: 24) indicates that only Gabon is an exception. However, both the PASEC microdata and information on the structure of Madagascar's schooling system point to grade 5 being tested there too.
[19] This was established through the relevant variable in the microdata.

PASEC is relatively well documented[20], though researchers have reported difficulties in obtaining the microdata, despite the fact that the website states these can be made available. For grade 6, proficiency is defined as having reached 'level 2', or at least a score of 518 in reading or 521 in mathematics. Importantly, for official SDG reporting, UIS uses a more stringent 'level 3' criterion[21]. While national end of primary statistics from PASEC align relatively well with non-PASEC statistics, such as the MICS numeracy statistics, the same cannot be said of PASEC's grade 2 statistics. The five countries common to PASEC and MICS correlate highly and positively across the programmes – this is using end of primary mathematics in PASEC. If grade 2 PASEC mathematics is considered, however, the MICS-PASEC correlation is low and slightly negative. For all 14 PASEC countries, correlations across the two levels are moreover low, whether reading or mathematics is considered. The conclusion that can be drawn is that PASEC grade 6 displays values one might expect, while the grade 2 patterns raise questions about reliability.

Importantly, the so-called Rosetta Stone study, published in 2022, drew from data where learners across three PASEC countries were tested using PASEC items, and items from TIMSS and PIRLS. This allowed a much more precise estimation of where in the PASEC scale the TIMSS- or PIRLS-based benchmarks prioritised by the UIS are situated. The TIMSS intermediate benchmark (475) was found to correspond to a score of 700 on the PASEC scale, while the relevant low benchmark in PIRLS of 400 corresponds to 640 on the PASEC scale[22]. These PASEC values are considerably higher than the abovementioned values of 521 and 518. If one adds the fact that the PASEC scores of 700 and 640 ought to be achieved in grade 4 already, then it becomes very clear that PASEC's end of primary proficiency threshold is very lenient, relative to the UIS standard. Arguably, alignment should be sought by using grade-on-grade gain assumptions to produce hypothetical PASEC grade 4 values, using PASEC grade 6 as the point of departure, and then locating, for instance, the equivalent of the TIMSS 475 score in the grade 4 distribution. However, there are margins of error associated with both these transformations whose impact would be difficult to ascertain. For this reason, the current report does not use the Rosetta Stone concordance tables, even if they are clearly of value for other analytical tasks[23].

A key advantage with PASEC is that it covers many African countries, that its level of difficulty is based on what skills learners in Africa typically display[24], and that the level of transparency, in terms of technical documentation and microdata availability, is relatively good. The Africa-only focus would however limit comparisons beyond the continent.

SACMEQ has assessed nationally representative samples of grade 6 learners and produced comparable measures of reading and mathematics across 15 mostly anglophone countries in Southern and East Africa since 2000. The values for the 13 countries in Table 1 are from the 2013 round of testing, the source being an official international SACMEQ report[25]. The language of the test is English, except for Portuguese in the case of Mozambique. To be considered proficient in Table 1, a learner had to reach at least level 5 of eight SACMEQ-defined competency levels. This threshold is based on the practice in the UIS SDG reporting system.

By 2021, when a new round of testing occurred, the acronym had changed to SEACMEQ. Proficiency statistics based on the 2021 testing for several countries are not readily available.

[20] CONFEMEN, 2017, 2022.
[21] See for instance CONFEMEN (2020: 75) and UNESCO Institute for Statistics (2021a).
[22] UNESCO Institute for Statistics, 2022b: 40-41.
[23] The PASEC-TIMSS linking within the Rosetta Stone initiative was used by Gust *et al* (2024: 6) for validation purposes.
[24] PASEC, but also SACMEQ, have displayed far less serious floor effects than the Latin American regional assessment programme LLECE – see Gustafsson and Barakat (2023).
[25] Awich, 2021: Tables A7a and A7d.

Individual countries are expected to produce country-level reports, but at the time of writing only the South African report seemed to be available online[26]. That report included 2021 mean values, but not proficiency statistics, for South Africa and several other countries. SACMEQ technical reports, especially concerning how comparability over time in learner results is achieved, are not readily available. Nor is the microdata from 2013 or 2021 available to researchers in general. This lack of transparency has raised questions around the reliability of the SACMEQ statistics. Yet published national statistics, at least insofar as they pertain to learner assessment results, appear plausible[27]. A comparison of the 2000 microdata of South Africa and Botswana, where the comparison was against TIMSS grade 8 results for the same countries, has been performed by Sandefur (2016). A comparison against TIMSS grade 4 results would have been relevant for the current analysis, but given that there are no explicit assumptions regarding the relationship between SDG end of primary and lower secondary proficiency statistics, the grade 8 TIMSS focus of Sandefur places that analysis outside the scope of the current paper.

Like PASEC, SACMEQ is a vital instrument gauging levels of proficiency among Africa's learners. It covers around as many countries as PASEC, has a somewhat longer history, and clearly much effort has gone into producing data which are as comparable as possible across countries. A key drawback is the low level of transparency in terms of the availability of microdata and metadata.

MICS is a major and well-documented UNICEF programme collecting data from nationally representative samples of households, with a focus on children's health and well-being. Since 2017, MICS surveys have included one-on-one assessment of children in the household, whether they attend school or not. The statistics on 18 African countries in Table 1 draw from data collected within the period 2017 to 2022 and are all reflected in the 2024 *The state of the world's children* report of UNICEF[28]. That source also includes a further 24 non-African countries. Three of the African countries had data collected at least partially in 2020 or 2021, meaning the pandemic is likely to have affected results: Benin, Eswatini, and Nigeria. Comoros had its data collected in 2022[29]. MICS assessed children in the age range 7 to 14. These statistics are repeated in voluminous MICS country-specific reports covering all aspects of MICS. In those reports, they are described as '4.1.1' statistics, meaning standards applicable to the relevant SDG proficiency indicators apply. The country reports make it clear that the overall subject-specific standards are 'successfully completed three foundational reading tasks' and 'successfully completed four foundational numeracy tasks'. The MICS reading assessment has three tasks: word recognition, 'literal questions', and a couple of 'inferential questions', all relating to a reading text. The numeracy assessment has four tasks: number reading, number discrimination, addition and pattern recognition[30]. This means in each subject children should successfully complete all tasks. As discussed above, how exactly to understand proficiency standards at the lower primary level is a matter where agreement between experts is still being sought. It is noteworthy that MICS-based proficiency values are considerably lower than what should be equivalent values in PASEC. To illustrate, for Madagascar MICS finds only 7% of children attending grades 2 or 3 proficient in mathematics, while PASEC finds 22% of grade 2

[26] South Africa: Department of Basic Education, 2024b.

[27] Rankings certainly appear plausible, though some trends over time are arguably questionable – see section 6. SACMEQ has also tested teachers, and here analysts have been more concerned around the reliability of the data and available statistics.

[28] See UNICEF (2014). An examination of the MICS website in mid-April 2025 indicated that no additional countries had new and finalised learning statistics, beyond what could be seen in the 2024 report.

[29] For Tunisia separate collections occurred in 2018 and 2023, but statistics in the table draw from the 2018 collection.

[30] UNICEF, 2020: 36.

children proficient in this subject. Such discrepancies are not surprising given the absence of finalised standards.

MICS's 18 countries in Table 1, many of which are not covered by other assessment programmes, makes MICS an essential source of information on learning proficiency in Africa, and for comparing proficiency in Africa to that in other world regions. As will be seen, for the purposes of the current paper a disadvantage with MICS is its 'ceiling effects', meaning its tests are set at a relatively low level of difficulty, thus precluding meaningful differentiation of proficient and non-proficient learners in grade 6. However, it will also be seen that that there are ways of working around this limitation.

PAL statistics provided for three countries in Table 1 are from a 2021 PAL report[31]. In all three countries, large nationally representative samples of over 100,000 children aged from 4 or 5 to 16 are covered, through household visits. Data collection was in 2015. Reading values are the 'proportion of children aged 14-16 able to read grade 2 level text', while mathematics values are 'proportion of children aged 14-16 able to solve Grade 2 level maths question'. Despite the labelling, the latter involves more than just one question, following general practice – see for instance PAL Network (2020). Yet, PAL results cannot be considered highly comparable, as standards are set nationally[32]. This should result in similar but not identical standards, and results which would be only loosely aligned to UNESCO standards. The languages used for the assessments are not made explicit in the main source document used here, though this aspect of the assessment is clearly important. From other PAL reports, it seems likely that English and Kiswahili were used. Technical documentation on the country-level PAL data collection processes is quite limited, and there seems to be no explicit emphasis on making microdata widely available to researchers.

EGRA in many ways pioneered one-on-one reading assessments of children, using methods such as words correct per minute (wcpm). It has been implemented in many languages. Despite the strength of the EGRA 'brand', what occurs under this label is not centrally controlled, and hence approaches can vary and comparability across countries and over time is not intended to be strong[33]. In fact, adaptation of the test to local contexts is encouraged[34], which would reduce comparability across countries.

The 21 EGRA reading results in Table 1 are taken from a spreadsheet accompanying the journal article Angrist *et al* (2021), an important contribution to efforts to harmonise learning data across countries[35]. In that analysis, the focus is on mean scores, not proficiency statistics. The final dataset spans 164 countries and the period 2000 to 2017. EGRA is used as a source for 71 of a total of 2,023 country-year data points. The 71 are from 48 developing countries across the world. The last available value per country in the Angrist *et al* source was taken, and then only countries in Africa were extracted, giving the 21 values of Table 1. Angrist *et al* clearly had access to and used the EGRA microdata, microdata which have been archived by USAID and which have never been readily available to researchers in general, even before the 2025 disruptions to USAID's work. Very importantly, EGRA testing, which takes place in schools, is very often not aimed at producing nationally representative samples. Schools targeted are often part of an intervention project, and specific to one region in a country. The Angrist *et al* spreadsheet has a column indicating whether the statistic is nationally representative. Nine of the 21 values in Table 1 are asterisked as they are not nationally representative. If older EGRA statistics in the spreadsheet, in other words not the most recent one, are used this does not change this picture at all. The frequent non-representativeness of EGRA is a key factor making

[31] PAL Network, 2021.
[32] PAL Network, 2021: 4.
[33] Gove *et al*, 2011: 29.
[34] USAID, 2016.
[35] The reference here is to both the 15-page article and a 25-page appendix.

this the least suitable source for the current paper of all the Table 1 sources. The fact that some EGRA statistics have been treated as if nationally representative when they are not in the final results of Angrist *et al* is obviously a limitation, which the authors acknowledge. Moreover, they attempt to reduce the across-comparability problem by drawing only from the reading comprehension part of the test, on the assumption that this suffers from fewer comparability problems across languages. In the absence of any other data for specific countries, EGRA could arguably be used cautiously to provide a sense of learning outcomes, especially if the sub-national sample seems relatively representative of the country.

EGRA is designed to test grades 2 to 4. Angrist *et al* use only data for these grades, but are not explicit about the extent to which certain countries tested a subset of the grades 2 to 4 range. However, they do draw from a sensitivity analysis which leads them to conclude that using grade as a predictor in calculating the Table 1 EGRA values is not necessary, as not doing this does not unduly influence the findings[36].

The 2025 disruptions to USAID have worsened the availability of EGRA metadata. An online data querying tool known as the Early Grade Reading Barometer, which included metadata reports, became unavailable in 2025. Even before 2025, however, reports describing the data collections behind many of the Table 1 statistics were not available online. Among other things, this reduces the chances of evaluating the usability of sub-national samples. As will be discussed below, fortunately some EGRA reports can be found on the web as they have been disseminated by organisations other than USAID.

The Early Grade Mathematics Assessment, or EGMA, was a USAID-funded assessment which complements EGRA. It is not considered in the current paper in part because it has been found to be even less comparable across countries than EGRA[37], and because it seems unlikely it would have brought additional value, beyond EGRA, to the current analysis.

2.3 Illustration of data coverage in Africa

The following two schematic maps of Africa[38] reflect how many values from Table 1 there are per country. The first map includes the least ideal source, namely EGRA, while the second map excludes EGRA. In the first, 47 countries, covering 97% of the continent's children[39], have data. Among the eight countries without data, representing the remaining 3%, around half of the children are in Somalia. In the second map there are measures of learning in 40 countries covering 79% of the continent's child population.

[36] See Patrinos and Angrist (2018).
[37] Cortez Ochoa and Sandoval Hernandez, 2023.
[38] This format is from UNESCO Institute for Statistics (2021c).
[39] Age 0 to 14 from UIS.Stat.

Figure 1: Number of values per country with EGRA

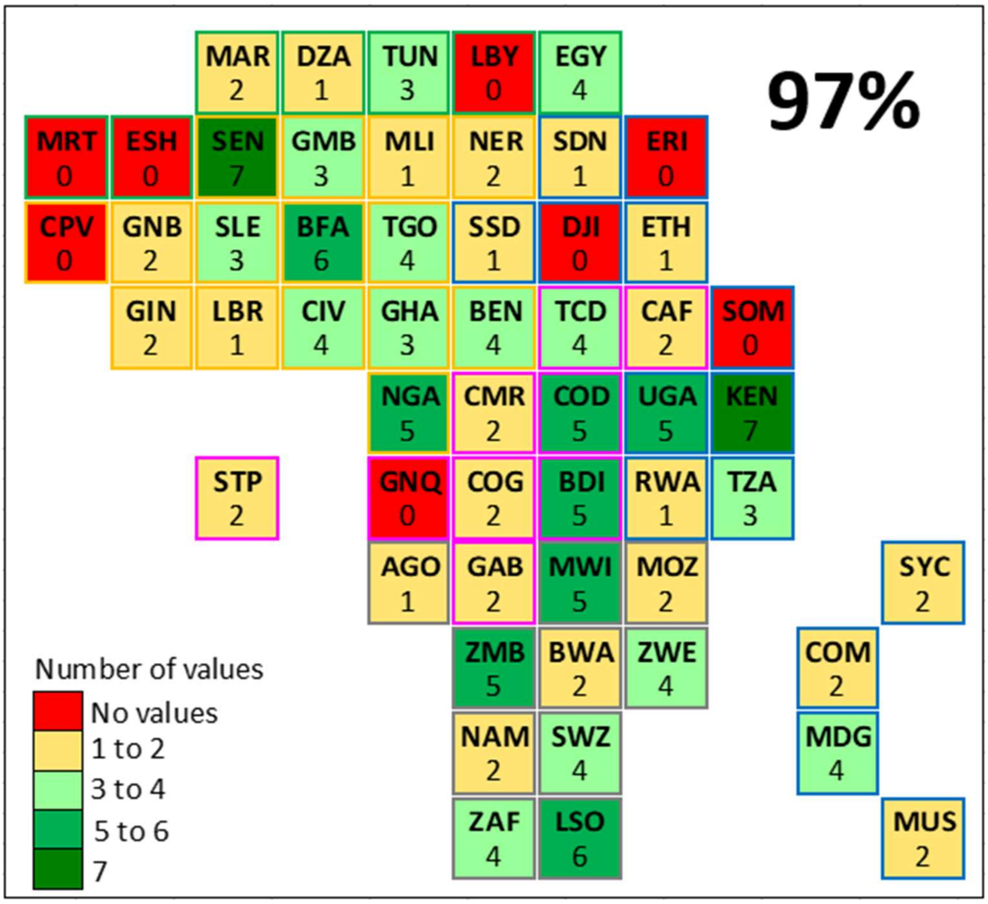
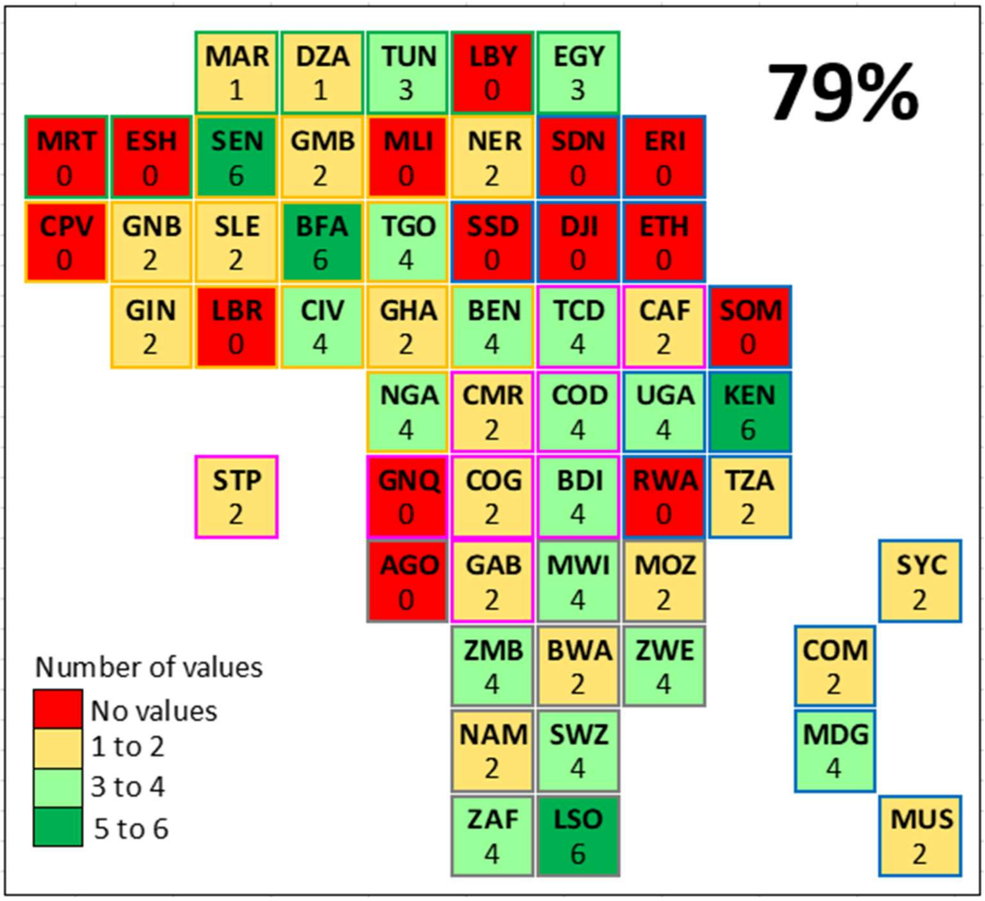


Figure 2: Number of proficiency statistics per country (no EGRA)



2.4 Confidence intervals around the proficiency statistics

Given the sample-based nature of the statistics from the nine programmes introduced in section 2.1, confidence intervals around the statistics should be considered. Standard errors, from which confidence intervals are derived, have been published for most proficiency values. Harmonisation can change confidence intervals. Specifically, in the typical approach where scores of individual learners are compared across two programmes in a linking country, confidence intervals found in each of the two programmes may compound each other, resulting in final confidence intervals which are wider than what was found in each of the two programmes originally[40]. Some sets of harmonised country statistics are presented without confidence intervals – this is the case in Gust *et al* (2024). Others do include standard errors – this is true for Angrist *et al* (2021).

In the interests of simplicity, the current paper avoids discussing in a step-by-step fashion the implications of the hybrid harmonisation approach pursued here for confidence intervals. The hybrid nature of the approach would make the calculation of final confidence intervals more complex than in the case of the typical microdata approach. For example, using results from one grade to estimate proficiency in another grade, something pursued below, could impact on standard errors in a manner that would be difficult to determine. However, confidence intervals around final statistics for a country are likely to be similar to those seen in the source programme. To assist in the interpretation of the final harmonised statistics, Table 3 below reports on 95% confidence intervals associated the proficiency statistics of specific programmes, as well as typical sample sizes. With regard to the latter, it is the number of

[40] Angrist *et al*, 2021: 11.

sampled schools or households which largely determines how narrow confidence intervals can be. The figures in the table draw from details relating to those countries for which the programme is the key data source (see section 3.7). Where a range of confidence intervals (CIs) is provided, the lowest would be for the smallest seen for any country for which the programme was the primary source of the final harmonised statistic, regardless of subject. Maximum CI values were derived similarly.

Table 3: Confidence intervals around available proficiency statistics

	Sample size (schools, unless otherwise indicated)	*95% CI – percentage points between lower and upper bound*
TIMSS[41]	±280	4 to 6
PIRLS[42]	192 to 321	5 to 8
LaNA[43]	±100	8
AMPL[44]	±250 to 290	2 to 9
PASEC[45]	180	3 to 7
SACMEQ[46]	126 to 295	3 to 4
MICS[47]	150 to 600 households	6 to 16
PAL[48]	68,588 households	±1
EGRA[49]		5 to 6

To illustrate the meaning of the CI values, a proficiency statistic of 30% in TIMSS would carry a confidence of interval of between 4 and 6 percentage points. The range exists largely because different countries have different sample sizes, but is also due to factors such as how

[41] Von Davier *et al*, 2024.
[42] Mullis *et al*, 2023.
[43] Von Davier *et al*, 2025.
[44] UNESCO Institute for Statistics, 2022a, 2024b. In the first of the two reports a bolding of Burundi's reading statistics, with the note that this means 'fewer than 30 students and/or 5 schools' were sampled, could be an error, as this is not consistent with other parts of the report (UNESCO Institute for Statistics, 2022a: 115). In this regard, it is also noteworthy that Burundi's reading proficiency statistic in AMPL is extremely low, and much lower than the statistic for mathematics – see Table 1 above.
[45] CONFEMEN, 2020: 32.
[46] Awich, 2021. Seychelles figures were not considered as a country this small would include all schools in the sample, have fewer tested schools than other countries (with samples) and thus, counter-intuitively, have large confidence intervals. Arguably, very small countries should be reported on differently, without applying standard statistical procedures, to prevent the impression that statistics are highly unreliable.
[47] Examination of a few national reports revealed that on average the number of households from which a child aged 7 to 14 was assessed was just over 5000. Moreover, these reports revealed that the confidence interval for the proficiency statistics of children in grade 2 or 3 was between 5 and 10. Analysis of the MICS microdata revealed that the number of sampled and tested children in grade 2 or 3 is roughly 1000 to 2000 per country, and for grade 6 between 150 and 600. Analysis of the data moreover pointed to a confidence interval for tested grade 6 learners of between 6 and 16. What appears in the table thus points to confidence intervals for proficiency statistics derived later in the report, and not for the Table 1 statistics, which refer to all children aged 7 to 14.
[48] Uwezo (2017) indicates that in 68,588 households across Tanzania (the only country for which PAL is the source in the final analysis) children aged 7 to 16 were tested. Assuming around 3 of 10 households among the 68,588 households had children aged 14 to 16, it can be concluded that roughly 20,000 households would be the source for Tanzania's results in Table 1 (the age range 7 to 16 is 10 age cohorts while 14 to 16 is 3 age cohorts). No details relating to standard errors could be found in the PAL reports, but some 20,000 households would result in a negligible confidence of interval of around just one percentage point.
[49] The Excel file accompanying Angrist *et al* (2021) has the standard error for the mean scores. Some examination of TIMSS statistics suggest that the standard error for the percentage proficient statistic would be just under half of the standard error for the mean. This resulted in the estimated 5 to 6 range shown in the table.

differentiated test booklets are used, or not, in the assessment. Assuming a CI of 6, we can be 95% certain that the true value lies in the range of 27% to 33%, the midpoint being the abovementioned 30% point statistic.

It is important to bear in mind that confidence intervals are not the only indication of how reliable specific proficiency statistics are. Other important factors would be the exclusion of certain categories of schools in some countries – see in particular the discussion in section 3.5.

3 Harmonising national statistics

3.1 The overall approach taken

It is instructive to examine the similarities and differences between the hybrid harmonisation process presented in the current paper, and typical harmonisation exercises in the past. Gust *et al* (2024) can be considered representative of the latter. That is also the most recent major work of its kind, and produced what is easily the most comprehensive set of harmonised learning measures for the world, largely because they resolved what had been persistent problems relating to measures for China and India.

Gust *et al* produced proficiency statistics for 159 countries relating to mathematics at the lower secondary level. Data on learning in science are permitted to influence the national statistics. The minimum standard used to define proficiency is essentially that of the SDG lower secondary proficiency indicator. Six international assessment programmes are used, and a requirement is that all must have available microdata. The microdata requirement means some data are relatively old – the oldest seems to be SACMEQ 2007 data[50]. The PISA[51] programme is used as the standard or anchor. Where 'linking countries' participate in PISA and another programme, the overlap is used to produce a translation of the scale of the other programme to the PISA scale. At least one linking country per other programme is required – in the case of the PISA-PASEC linking the fulfilment of this condition relies wholly on just one linking country, Senegal. Though data such as PASEC do not represent the secondary level, such data are deemed useful as country rankings at the primary and secondary levels are likely to be very similar. Statistics for China and India, who jointly account for a third of the world's children, are calculated through combining data from two sources: a collection in just a part of each country through one of the six international programmes, and a national programme that assessed learners across the whole country. Previous harmonisations had relied purely on international assessment data from very limited sub-national regions of these two countries. The end result is that for 85% of the world's population national statistics based on assessment data were produced, while for a further 13% learning outcome values were imputed from data on other variables such as income and enrolment rates. While 85% of the world's population is covered by the assessment data used, this drops to just 50% in the case of Africa, where assessment data from 29 of 55 countries were available[52]. Lastly, adjustments to each country's proficiency statistic to take into account those not in school were undertaken. Gust *et al* estimate that around a third of young people around the world do not receive secondary schooling. That fraction would clearly be lower for the primary level.

The hybrid harmonisation presented below responds to a somewhat different need to that of, for instance, Gust *et al*. The latter is strongly focussed on producing a global aggregate statistic, and global patterns that can inform analysis of, for instance, the impact of changing levels of proficiency across the world on subsequent economic growth. The aim below is more to inform discussions in a region, namely Africa, around the distribution and depth of 'learning poverty',

[50] For PASEC, 2014 data were used.
[51] Programme for International Student Assessment.
[52] If imputed values are considered, this 50% rises to 97%, reflecting the particularly strong dependence on imputations for Africa.

how to resolve this, and also how to improve the monitoring of the situation. To a significant degree, the audience is assumed to be national researchers and policymakers in the education field with a specific interest in their country. This explains why the emphasis is on drawing from whatever proficiency statistics exist, even if the underlying microdata are not accessible, evaluating the reliability of different statistics, where more than one per country exists, and on avoiding the imputation of proficiency statistics in the few cases where there were no usable educational quality statistics. Regarding the latter, filling in the gaps through imputation could detract from the importance of ensuring that the gap is filled through the required education data. The less stringent data requirements for this paper's hybrid harmonisation mean that relatively few countries are left with no educational data at all from which information can be drawn. As indicated earlier, such countries come to 3% of African children, compared to 15% of the world (and 50% of Africa) in the case of Gust *et al*. The wider variety of sources used for the current work results in instances where inconsistencies in the ranking of countries are brought to the fore. This, in turn, brings about the need for informed judgements around which source to prefer. Such decisions are largely absent in the earlier harmonisation analyses.

Though Angrist *et al* (2021) focus on producing mean scores per country across the world, and not proficiency statistics, it is in some ways more similar to the current exercise than Gust *et al* (2024). Above all, through extensive use of EGRA, but not MICS, Angrist *et al* are able to cover 94% of Africa's children with respect to primary level mean reading scores – Figure 1 points to a similarly high 97% level of coverage, though the intention in the current paper is to prefer MICS over EGRA[53]. Angrist *et al* do not estimate educational quality values using non-assessment data, as in the case of Gust *et al* and several other analysts. In Angrist *et al*, the most recent Africa data are from 2017. According to Table 2, around two-thirds of the African countries analysed for the current paper have a data source that is more recent than 2017. A comparison against the final figures of Angrist *et al*, is presented in section 3.7.

For the harmonisation presented below, adjusting statistics of countries not testing the benchmark grade is considered important, in line with the assumption that the audience is in part those with an interest in the comparability of the results of specific countries. Such adjustments are not used in Gust *et al* or earlier harmonisation exercises, it would appear, where the emphasis is to a fair degree on across-country averages, as opposed to findings on individual countries[54]. What follows makes some use of microdata, though where microdata were not available harmonisation occurred only through the comparison of national statistics. Without microdata, aligning statistics across two programmes requires at least two linking countries, preferably at least one near the bottom of the performance distribution and another near the top. In theory, this requirement can be relaxed if relationships between each programme and all other programmes are considered simultaneously[55], though such an approach did not appear necessary here.

The hybrid approach followed here produces statistics with less reliability than those derived from a process where the microdata of programmes are compared. The advantage with the current approach is that it lends itself more to a discussion of the various sources of learning data, and of how improvements in the monitoring of learning are unfolding.

The principles of any harmonisation process are only really useful to the extent that overlaps across programmes, or linking countries, permit their application. Table 4, which has a marker wherever statistics exist, suggests the linking country situation in Africa is relatively good. Four programmes, highlighted in grey, have been considered for the purposes of this paper to

[53] Angrist *et al*'s 97% drops to 61% if EGRA is discounted as a source.
[54] For instance, Gust *et al* ignore the fact that South Africa and Botswana tested grade 9 in TIMSS, and not the standard grade 8, something which would make the results of these countries appear artificially high.
[55] Gustafsson, 2013.

represent a global standard relatively well. AMPL is explicitly designed to do this, while the three IEA programmes of TIMSS, PIRLS and LaNA are especially good at advancing comparability among themselves and across different regions of the world. There are 13 countries participating in these standard programmes. Of the 13, four countries have values for more than one standard programme, allowing for some testing of the assumption that the four standard programmes successfully follow the standard. The four countries are: South Africa, Egypt, Burkina Faso, and Senegal. Four countries are present in PASEC and in at least one of the four standard programmes. These are marked in green. The absence of the very lowest performing PASEC countries in this group of four is potentially a problem. SACMEQ also has four countries participating in at least one of the standard programmes. The four SACMEQ linking countries are marked in blue. Red markers for MICS and EGRA represent countries with a link to the standard programme. There are only two linking countries of this category in MICS suggesting that additional mechanisms should be explored to align MICS values. In the case of EGRA, there are seven linking countries.

Only seven countries would need to draw from MICS data, assuming that the aforementioned programmes are preferable. The seven are marked by borders in the MICS column. For one country, Tanzania, PAL emerges as the best available measure. Lastly, for a further seven countries one would need to resort to EGRA data, clearly the least comparable of the nine programmes – see the countries in rectangular borders in the EGRA column. Of these seven, four have statistics which have been found not to be nationally representative in Angrist *et al*: Ethiopia, Liberia, Mali and South Sudan. Hollow markers mean the EGRA statistic is sub-national.

Table 4: Initial evaluation of linking countries

	MICS	PASEC	SACMEQ	LaNA	TIMSS	PAL	AMPL	PIRLS	EGRA
AGO									●
BDI		●					●		●
BEN	●	●							
BFA		●		●			●		
BWA			●						
CAF	●								
CIV		●					●		
CMR		●							
COD	●	●							○
COG		●							
COM	●								
DZA					●				
EGY				●				●	●
ETH									○
GAB		●							
GHA	●								●
GIN		●							
GMB	●								●
GNB	●								
KEN			●			●	●		○
LBR									○
LSO	●		●				●		
MAR					●				○
MDG	●	●							
MLI									○
MOZ			●						
MUS			●						
MWI	●		●						●
NAM			●						
NER		●							
NGA	●			●					○
RWA									●
SDN									●
SEN		●		●			●		●
SLE	●								●
SSD									○
STP	●								
SWZ	●		●						
SYC			●						
TCD	●	●							
TGO	●	●							
TUN	●				●				
TZA						●			●
UGA			●			●			○
ZAF			●		●			●	
ZMB			●				●		●
ZWE	●		●						

3.2 Informing grade-on-grade gain assumptions

The descriptions of the assessment programmes provided in section 2 indicate that of the six programmes other than EGRA where assessment occurs in schools, four reflect a situation where different grades are tested in different countries. The two exceptions are PIRLS and SACMEQ. The four programmes in question and their details are shown in Table 5 below. In 18 instances testing did occur in the standard grade – this would be the sum of the 'No adjustment' column. As discussed previously, the standard grade in TIMSS and LaNA would be grade 4, while in PASEC it is grade 6. For AMPL grade 6 has been selected as the standard for the current paper. In eight instances a non-standard grade was tested. In TIMSS, South Africa (ZAF) tested grade 5 instead of the standard grade 4, thus an adjustment would mean estimating proficiency one grade down. In LaNA, countries tested either grade 5 or 6, instead of the standard grade 4, meaning that for this programme downward adjustments would be needed for all four countries. In AMPL there is data for grade 7 in the case of Zambia, necessitating a downward adjustment. Finally, in PASEC two countries tested grade 5, meaning upward adjustments would be needed here.

Table 5: Grade anomalies in programmes assessing schools

Programme	*No adjustment*	*Move one up*	*Move one down*	*Move two down*
TIMSS	3		ZAF	
LaNA	0		EGY	BFA, NGA, SEN
AMPL	5		ZMB	
PASEC	12	GAB, MDG		

Grade anomalies for EGRA will be discussed in section 3.6, when it becomes clearer for which countries EGRA is an unavoidable source.

The current paper assumes that grade 6 should be compared across countries, even if the SDG indicator in question refers to 'end of primary'. It is thus assumed that it is unfair to compare, say, Zambia's grade 7 proficiency to Côte d'Ivoire's grade 6 proficiency, even if in both cases this represents the end of primary schooling. It would be unfair because Zambia's learners have enjoyed the benefit of one extra year of schooling. However, it is acknowledged that this assumption is debatable. School curricula are likely to be designed in a manner that creates similar expectations across countries with respect to the end of primary, even if different grades apply. Yet the assumption that the same grade should be compared across countries remains a compelling one[56].

A few analyses exist that estimate grade-on-grade gains in different schooling systems. Typically, these gains are expressed as a proportion of a standard deviation. A seminal text in this area is Hill *et al* (2008), who use United States data to conclude that gains decline from around 1.00 standard deviation for grades 1 to 2 to around 0.35 standard deviations for grades 5 to 6. Reading and mathematics gains are reported separately, but differences across subjects are small. In developing countries grade-on-grade gains can be expected to be considerably lower than this.

The last column of Table 6 provides a few Africa-specific findings in relation to grade-on-grade progress in standard deviations. Singh (2019) finds one grade to produce progress coming to between 0.35 to 0.40 standard deviations in mathematics (M) in the initial grades. South Africa's national assessment points to annual progress in the range grade 3 to 6 in reading (R) coming to 0.35 standard deviations. This is one-third of the 1.05 gain seen between grade 3 and grade 6, so across all three grades. The corresponding grades 3 to 6 mathematics gain is notably high. Thereafter MICS microdata corresponding to four of the mathematics statistics for MICS seen in Table 1 are used. With the exception of DRC's very low grade 5 to 6 gain[57], the gains seen in MICS are not that different to gains seen in other data.

[56] A further complexity arises when a country introduces a grade below grade 1 in primary schools, which in effect becomes the first grade of primary schooling, making grade 1 in effect the second grade. South Africa's recent universalisation of 'grade R' in schools, below grade 1, is a case in point.
[57] The reason for the small gain in DRC is not a result of a small sample. N for grades 5 and 6 learners is 1344 in the DRC microdata.

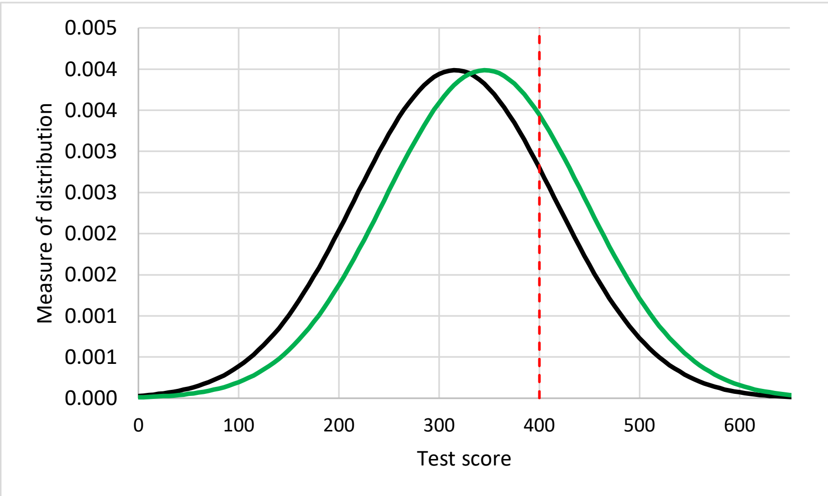
Table 6: Grade-on-grade gains in developing countries

Country	*Grades*	*Reference*	*Annual gain as s.d.*
Ethiopia	Initial	Singh, 2019: 1785.	M: 0.35 to 0.40
South Africa	3 to 6	South Africa: Department of Basic Education, 2024: 40-1.	R: 0.35 M: 0.47
	6 to 9		R: 0.30 M: 0.29
Madagascar	4 to 5	Analysis of MICS microdata.	M: 0.21
Benin	5 to 6	Analysis of MICS microdata.	M: 0.32
Chad	5 to 6	Analysis of MICS microdata.	M: 0.30
DRC	5 to 6	Analysis of MICS microdata.	M: 0.10

If a country's grade 5 proficiency statistic must be raised so that it reflects what is likely in grade 6, how would the above standard deviations be used? First, an assumed magnitude for the gain would have to chosen. Let us assume this is 0.30, based on the above table. The following formula, which uses Excel functions, produces the new proficiency estimate. The new proficiency statistic P_2 is a function of three other variables only: the original proficiency value P_1, the assumed gain in standard deviations G, and the number of grades for which an adjustment must be made N. If G is negative this means there must be a subtraction because the proficiency statistic must be reduced. Here the value 100 is the assumed standard deviation in the initial distribution, while 400 is the minimum proficiency level required. These two values can be changed to any plausible value. P_2 would not change. To illustrate a result, if the initial level of proficiency P_1 is 20%, if the gain in standard deviations is 0.30, and an adjustment of just one grade is needed, then the new level of proficiency P_2 is 29.4%. This change is illustrated graphically in Figure 3. The normal distribution moves to the right by 30 points, which is 0.30 of the standard deviation of 100. The area below the curve and to the right of the 400 threshold grows from 20% to 29.4%.

$$P_2 = 1 - NORMDIST(400, (400 + 100 * -NORMINV(1 - P_1, 0,1)) + G * N * 100,100,1) \qquad (1)$$

Figure 3: Theoretical distribution of scores before and after a .30 s.d. increase



3.3 Step 1: Within-programme adjustments

The only adjustments that seemed necessary for individual programmes, before comparisons between them commenced, was grade-related adjustments, using the approach outlined in section 3.2. Adjusted values appear in the last column of Table 7. The default assumed standard deviation change for one grade was 0.30. Exceptions were made for Madagascar, for which a relatively low value of .21 was seen in Table 6, and for South Africa, whose Table 6 figures suggested that mathematics grade-on-grade gains were likely to be somewhat greater than 0.30.

Table 7: Grade-related adjustments

Country	*Programme*	*Required grade adjustment*	*Initial %*	*Gain in s.d.*	*New %*
Reading					
EGY	LaNA	-1	42.0	-.30	30.8
BFA	LaNA	-2	24.0	-.30	15.7
NGA	LaNA	-2	20.0	-.30	12.7
SEN	LaNA	-2	12.0	-.30	7.0
ZMB	AMPL	-1	9.7	-.30	5.5
GAB	PASEC	+1	93.4	.30	96.5
MDG	PASEC	+1	17.5	.21	23.4
Mathematics					
EGY	LaNA	-1	29.0	-.30	19.7
BFA	LaNA	-2	13.0	-.30	7.7
NGA	LaNA	-2	11.0	-.30	6.3
SEN	LaNA	-2	8.0	-.30	4.4
ZAF	TIMSS	-1	17.0	-.35	9.6
ZMB	AMPL	-1	16.0	-.30	9.8
GAB	PASEC	+1	66.7	.30	76.8
MDG	PASEC	+1	21.6	.21	28.2

3.4 Step 2: Removal of implausible values

In this step the aim was to identify country values which emerged as outliers when programmes were compared, with a view to excluding such country values, especially if either the country was not a linking a country, or if a country value had apparently a more reliable alternative elsewhere in the data.

The approach followed was to first take the 15 adjusted values from the last column of Table 7 and to insert them as replacements into Table 1. Thereafter, every column in Table 1 was compared to every other column. Specifically, the values from the first were regressed on the values of the second wherever the two had at least three countries in common. This allowed for predicted values in the base, or dependent, column to be calculated. The standard deviation of the original dependent values was found. Then the difference between the actual and predicted values in the dependent column were calculated. If that difference divided by the standard deviation was greater than a particular threshold, then the country value was considered an outlier. Basically, outlier status was identified if a country's value was higher or lower than the predictor values would suggest. The threshold used was 0.95 of a standard deviation, which produced 20 outlier values. If the threshold had been zero, in other words if perfect alignment had been the expectation, then 432 country values would have been considered outliers. The 0.95 threshold thus made around 5% of the 432 emerge as outliers.

The 20 outliers are shown in Table 8. For example, when MICS reading was regressed on PASEC reading, a regression involving five countries, Chad's (TCD) actual MICS value was considerably lower than its predicted value, and Madagascar's MICS value was considerably higher than its predicted value. Low outliers are marked in red in the table.

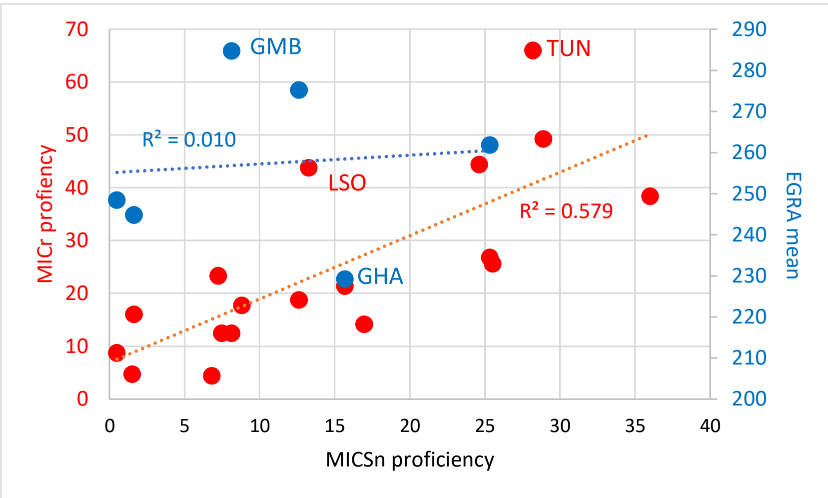
How important for the harmonisation are the 20 values implied by Table 8? According to Table 4 above, Nigeria and Lesotho are important for linking MICS, and unfortunately both these countries emerge as outliers within MICS, whether MICS reading or numeracy is considered. This strengthens the need to have additional or alternative linking countries for MICS. Outliers in EGRA – the last row of the table – are not surprising given the EGRA issues. Here it seems prudent to remove from the subsequent harmonisation steps three EGRA countries: Ghana, Gambia and DRC (COD). These three countries, according to Table 4, do not rely on EGRA as a last resort for an educational measure, neither are they EGRA linking countries. Even Burundi, which is a linking country, could be removed given that EGRA has several other linking countries.

Table 8: Outliers when comparing predicted to actual

MICSr		LSO TUN	TCD MDG		COD NGA
MICSn	LSO STP				COD NGA
AMPLr				BDI SEN	
AMPLm					BDI SEN
EGRA	GHA GMB	GHA GMB	COD BDI		
↑ Dependent Predictor →	MICSr	MICSn	PASECr	PASECm	EGRA

The following graph illustrates a few of the outliers referred to in Table 8. The very poor alignment of EGRA (right-hand vertical axis) with, in this case, MICSn is clear. Given the low correlation, the finding that Gambia (GMB) and Ghana (GHA) are outliers does not mean much. The pertinent finding is that *no* EGRA statistics appear to align well[58]. The regression of MICS reading on MICS mathematics is also shown for illustrative purposes.

Figure 4: Outliers with MICSn prediction



Note: Red markers should be read against the left-hand vertical axis, and blue markers against the right-hand axis.

3.5 Step 3: Examination of values reflecting the standard

Figure 5 reflects values for 13 countries participating in at least one of the four standard programmes, after the grade-related adjustments have been applied. A few striking anomalies clearly exist. Most striking is perhaps Senegal's (SEN) mathematics proficiency of 4% in LaNA against 34% in AMPL. Even before Senegal's grade-related downward adjustment in LaNA, proficiency in mathematics was a low 9%, so the anomaly is not primarily an outcome of this adjustment. One important factor that influences the Senegal anomaly is that LaNA always produces higher values for reading than for mathematics, while the reverse applies for AMPL (if one ignores the exception of Côte d'Ivoire). This must reflect differing interpretations of the

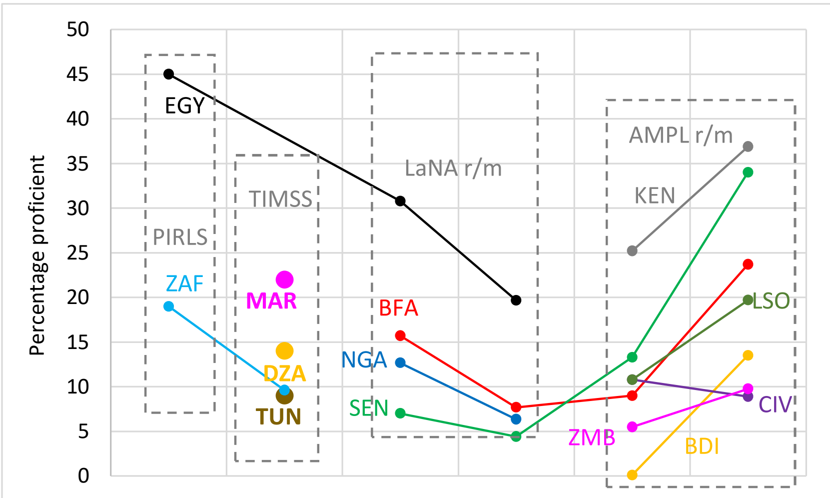
[58] One might expect 'mirror comparisons', such as MICSr on EGRA and EGRA on MICSr to produce the same outliers, but with opposite signs. This does not have to be the case, and is not the case in many instances in Table 8. While the R squared would be the same, outlier countries would differ as the vertical misalignment would refer to a different programme in each of the two analyses.

UNESCO standards, and suggests there is no clear reason for the harmonisation to end up with reading values being higher than mathematics values, or mathematics values being higher than reading values.

In the one instance where country rankings can be compared across programmes, there is less consistency than might be expected: in LaNA the values for Burkina Faso are always higher than those for Senegal, while the reverse is true in AMPL. Ranking anomalies are clearly more concerning than value anomalies. Programmes may interpret standards differently, but one might expect rankings to remain consistent. Inconsistent rankings are likely to be the result of sampling problems. It was not possible to compare both Senegal and Burkina Faso across the microdata of both LaNA and AMPL, as in the case of the latter programme the two countries were not present[59]. However, comparing the LaNA data for the two countries against World Development Indicator values pointed to electricity access in the Burkina Faso LaNA sample being too high[60], suggesting less advantaged learners were under-represented[61]. This would be consistent with the hypothesis of Burkina Faso's proficiency values being over-estimated in LaNA.

While the picture in Figure 5 undoubtedly reflects the fact that there is considerable room for better standardised measurement of learning, this picture is not surprising. As discussed below, comparing TIMSS and PIRLS, both IEA programmes, reveals that rankings can easily change.

Figure 5: Four programmes considered the standard

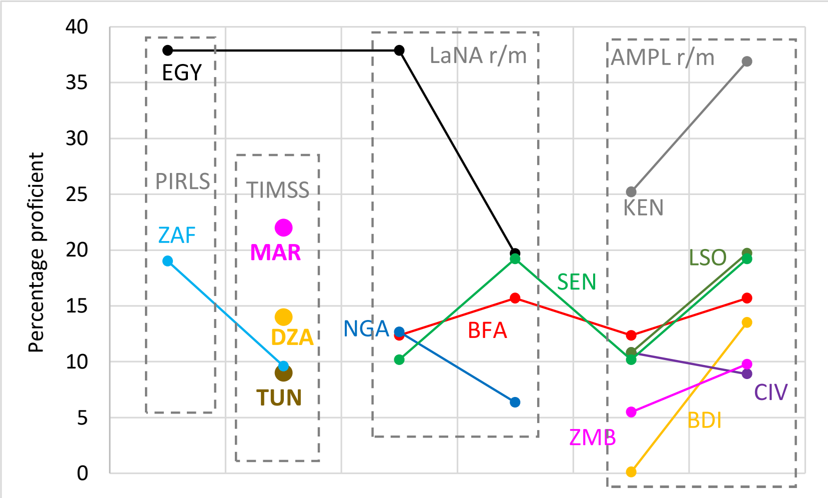


Note: For LaNA and AMPL the reading values are represented on the left, the mathematics values on the right.

Figure 6 provides a minimalistic harmonisation where differences across subjects are allowed to persist. The only change between this picture and the previous one is, firstly, that for Burkina Faso and Senegal, averages for each subject have been calculated and, secondly, that Egypt's reading values in PIRLS and LaNA have been changed to the average across the two. The result is that no country in the graph displays more than one reading value and no country displays more than one mathematics value.

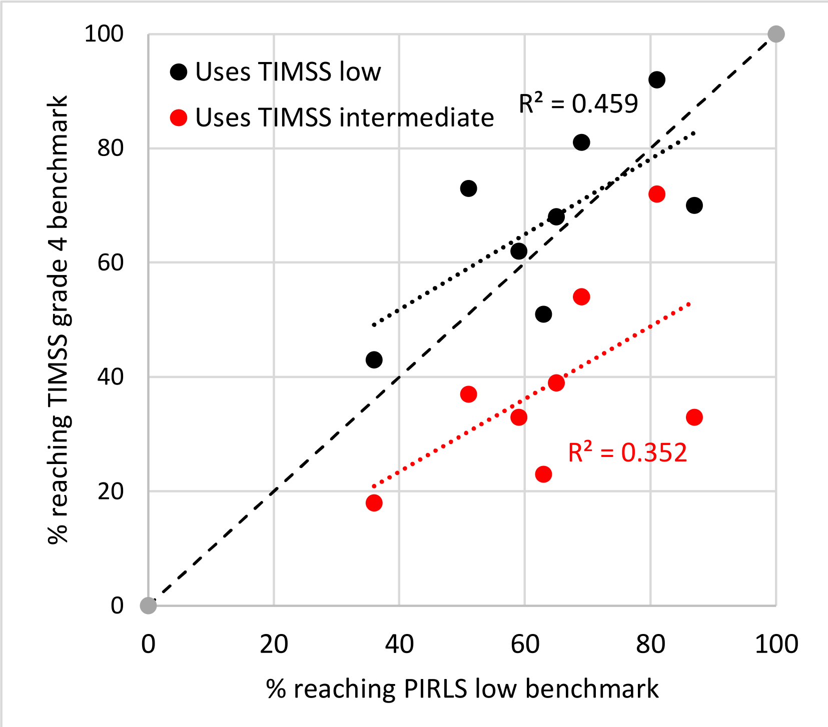
[59] Data for Gambia, Kenya, Lesotho and Zambia were available at https://ampl.uis.unesco.org/amplab-databases-and-analysis (July 2025).
[60] Specifically, in LaNA 47% of learners had access to electricity in the home (Von Davier *et al*, 2025: 34), against 22% in the World Development Indicators of the World Bank (https://databank.worldbank.org/source/world-development-indicators). In contrast, the corresponding Senegal figures of 81% and 74% are considerably closer.
[61] Von Davier *et al* (2025) indicate that 5.5% of schools were excluded from the sample, mainly private Muslim schools. The electricity figures suggest the exclusions may have been more extensive.

Figure 6: Harmonisation of standard values with subject differentiation



A clear problem with the above representation is that three North African countries – Morocco, Algeria (DZA) and Tunisia – appear to have proficiency values which are too low. These are countries with above average per capita income within the continent, yet in the graph their educational quality is similar to several countries with much lower levels of economic development. There is a simple explanation for this anomaly. The three are rated on the basis of their mathematics proficiency only, yet their reading values can be expected to be high given the very large across-subject proficiency differences in the well-established TIMSS-PIRLS system of the IEA. The graph illustrates this difference in the case of South Africa (ZAF). Figure 7 illustrates this difference for more countries, specifically for any of the bottom ten performing countries in PIRLS 2016 which had TIMSS 2019 grade 4 mathematics values, or any of the bottom ten performing countries in TIMSS 2019 grade 4 mathematics which had PIRLS 2016 values. The average difference between the relevant PIRLS and TIMSS proficiency statistics for the eight countries in the graph is 33 percentage points. In AMPL and LaNA, based on Figure 5, this difference is 10 points in both cases (though the highest values are for a different subject across the two programmes). Thus, not having the presence of a reading, specifically a PIRLS value, disadvantages the three countries in the rankings.

Figure 7: Relationship between PIRLS and TIMSS proficiency values



Source: Mullis et al, *2017, 2020.*

The use of the low TIMSS benchmark is illustrated in Figure 7 as background information. Clearly, had this benchmark been used, reading and mathematics proficiency values would have been more similar. It is noteworthy that for the eight countries represented by the red markers the ranking differs considerably depending on whether the horizontal or vertical axis is used. The rankings of seven of the eight countries change, the average shift in ranking being 1.6.

Figure 8 reflects the PIRLS-TIMSS average proficiency, using the intermediate TIMSS benchmark, as a function of the TIMSS intermediate value, using the same eight countries as for the previous graph. This function allows for reading-mathematics averages to be estimated for the three North African countries.

Figure 8: Relationship between TIMSS value and the two-subject mean

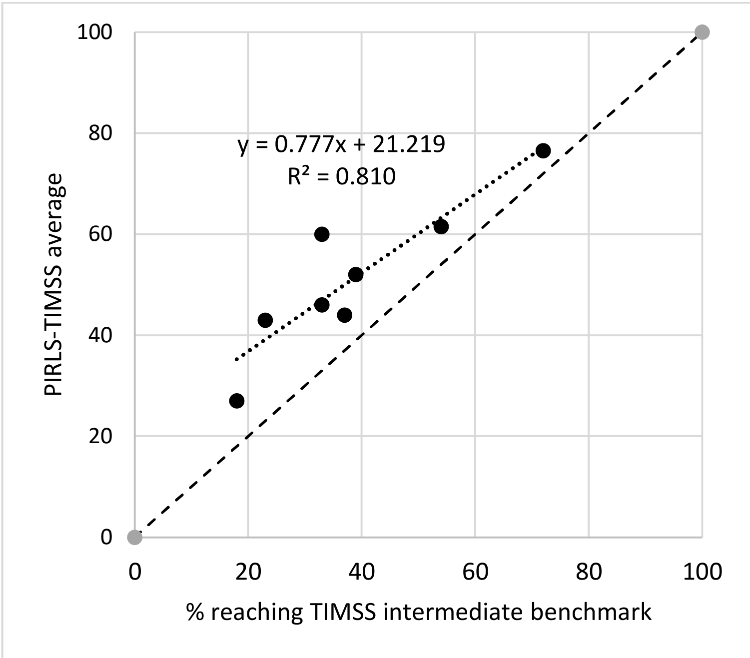
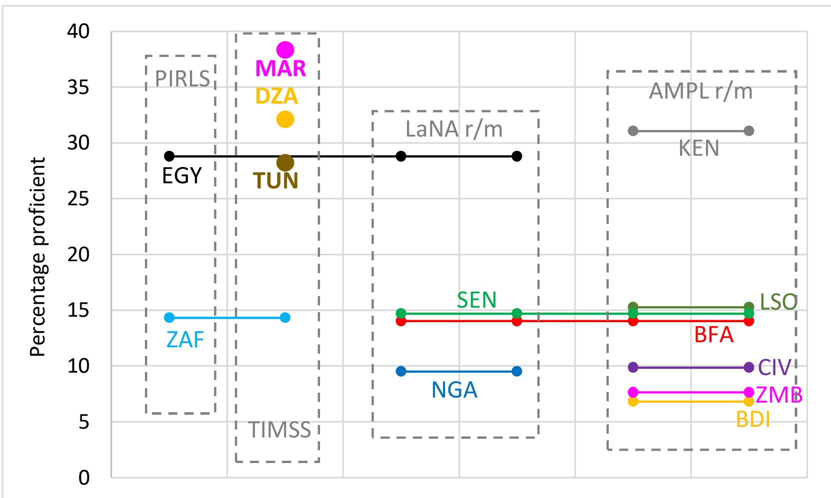


Figure 9 below uses the estimated across-subject averages for Morocco, Algeria and Tunisia. It also reflects the across-subject averages for the other ten countries. The advantage with across-subject averages seems twofold. Firstly, this allows for necessary adjustments where a country only has a value for one of the two subjects. Secondly, it eliminates much of the 'noise' created by often inconsistent across-subject differences. With the noise removal, certain patterns become clearer, such as that Kenya's learning levels appear to be considerably higher than those seen in the rest of Sub-Saharan Africa.

Economists and analysts looking at development in general tend not to be too concerned with subject-specific statistics. What tends to be important for them are general measures of learning, as in Figure 9. On the other hand, education planners tend not to work with these kinds of averages, and prefer breakdowns by subject, as in Figure 6. The remainder of this paper will attend to both needs by focussing both on subject-specific results, as well as a general measures of learning per country.

Figure 9: Harmonisation of standard values without subject differentiation

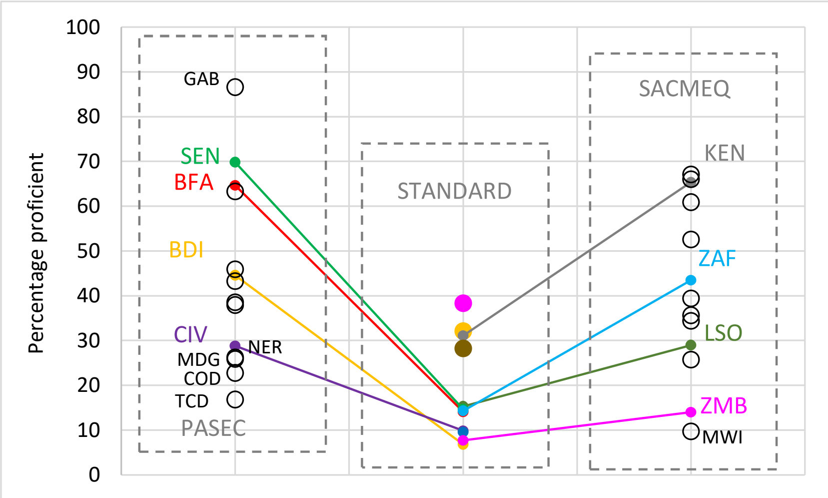


3.6 Step 4: Iterative alignment of non-standard values

3.6.1 PASEC and SACMEQ

Figure 10 below illustrates that there are four countries linking PASEC countries to the set of standard values seen in Figure 9 – the colouring for the standard values has remained the same – and also that there are four countries linking SACMEQ to the standard. The focus here is on averages across the two subjects. It is moreover clear that for linking countries PASEC and SACMEQ values are considerably higher than those in the standard.

Figure 10: PASEC and SACMEQ before adjustments without subject differentiation



The alignment of SACMEQ values from 2013 had to occur without any use of microdata, as the latter were not available. The reading and mathematics analysis of Figure 11 draws from the original SACMEQ proficiency values of Table 1, in relation to the horizontal axis, and from the standard subject-specific values for the four linking countries seen in Figure 6, in relation to the vertical axis. The intercepts and coefficients from this analysis are used to transform all SACMEQ reading and mathematics values, producing the values seen in Table 9. For the across-subject averages, the horizontal axis represents the average across the original SACMEQ values, while the vertical represents the standard values from Figure 9. This produced more plausible patterns than not having a separate transformation for the average, and simply averaging across the two transformed subject-specific values. The transformed average values are illustrated in Figure 13 below. Fortunately, in the case of SACMEQ the linking countries cover both the bottom of the range, through Zambia, and the top of the range, through Kenya. This reduces the seriousness of the SACMEQ microdata not being available, though obviously having the data would have improved the reliability of the harmonisation somewhat.

Figure 11: Standard value predicted by SACMEQ

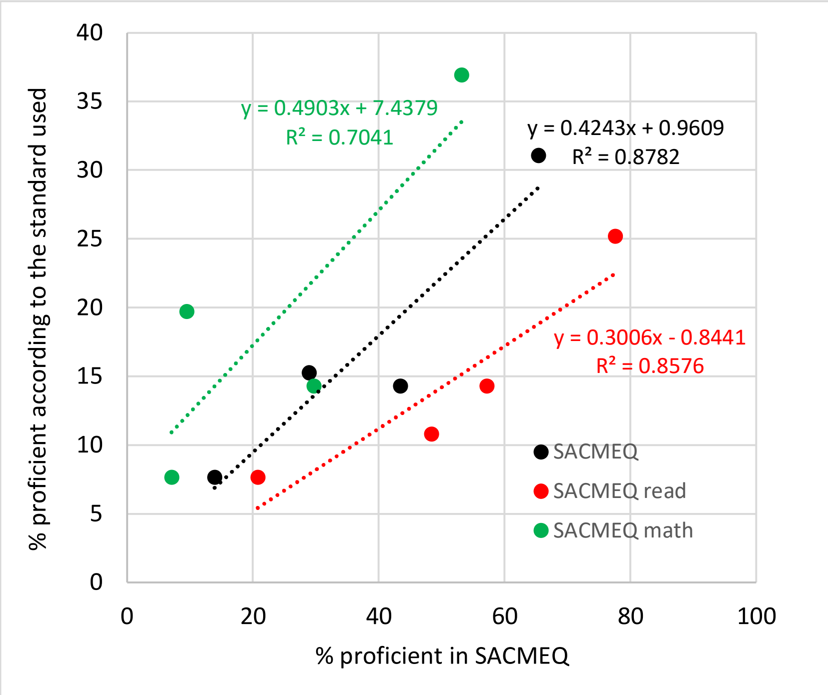


Table 9: Adjusted SACMEQ values

	Reading	*Mathematics*	*Average*
Intercept	-0.8441	7.4379	0.9273
Slope	0.3006	0.4903	0.2271
BWA	19.7	25.4	21.4
KEN	22.5	33.5	26.7
LSO	13.7	12.1	11.8
MOZ	10.1	14.9	10.5
MUS	21.8	36.3	27.4
MWI	3.8	9.4	4.0
NAM	17.6	16.0	16.1
SWZ	24.5	25.8	24.8
SYC	23.1	33.1	26.9
UGA	14.4	17.6	14.5
ZAF	16.4	22.0	17.7
ZMB	5.4	10.9	5.7
ZWE	12.8	18.9	14.0

It should be noted that the approach taken with respect to SACMEQ results in a situation where although reading results are always higher than mathematics results using the original Table 1 values, the end result of Table 9 is that most countries display better mathematics than reading results. This is because three of the four SACMEQ countries which can be linked to the standard display better mathematics than reading results within the standard – see Figure 6.

PASEC linking countries are clearly less representative of the full range of PASEC countries. Fortunately, PASEC 2019 data were available to guide the harmonisation. Also fortunate is the fact that LaNA microdata became available shortly before this work was undertaken. This allowed for so-called concordance tables, to use the terminology of the Rosetta Stone studies. Two countries were common across PASEC and LaNA: Senegal and Burkina Faso. As shown in Table 7, reading proficiency according to LaNA was 7.0% and 15.7% for the two countries respectively, after grade-related adjustments. Put differently, 7.0% and 15.7% of learners obtained 400 points or above on the LaNA scale. One might expect 7.0% of Senegal's learners and 15.7% of Burkina Faso's learners to achieve in excess of the same specific PASEC threshold score. However, for various reasons, including even slight inconsistences in the sampling approaches and the nature of the tests, this ideal is unlikely to be seen in the real world. Indeed, in the Senegal PASEC data 7.0% of learners obtained a score of 708 or higher, while 15.7% of Burkina Faso's learners obtained a score of 639 or higher. The higher threshold in Senegal, when in theory the threshold across the two countries should be same, is consistent with Senegal performing slightly better than Burkina Faso in PASEC in terms of the PASEC proficiency level for reading – see Table 1.

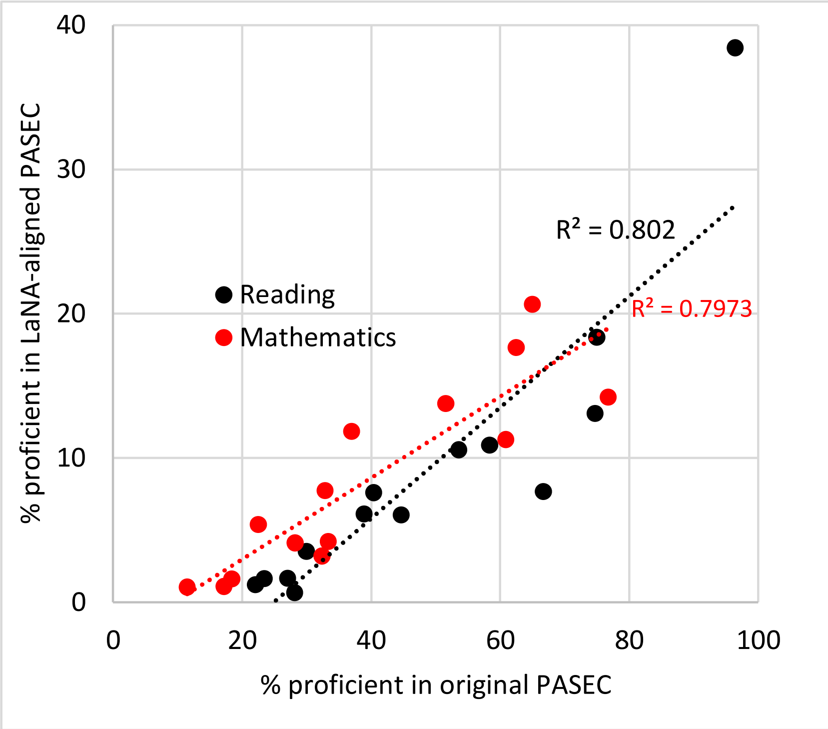
The data thus suggested that the relevant threshold for reading in PASEC in order to LaNA-align the PASEC values lay between 639 and 708 PASEC points. After some interrogation of the data, using options within the 639 to 708 range, a score of 673 was selected as this maximised the alignment across the four linking countries present in PASEC and the standard[62]: Burkina Faso, Senegal, Burundi and Côte d'Ivoire. The resultant LaNA-aligned reading proficiency statistics for all PASEC countries, based on the microdata, are shown in Table 10[63].

The situation was less straightforward for mathematics. The PASEC-LaNA concordance tables suggested the proficiency threshold should lie in the range of 671 and 717 PASEC points, these being the findings for the Burkina Faso and Senegal data respectively. However, even the lowest value in this range, 671, produced proficiency statistics which were implausibly low for the full set of PASEC countries. This is probably because of measurement idiosyncrasies across the two programmes that were specific to Senegal and Burkina Faso. In the end, the PASEC threshold which minimised differences across the standard and PASEC with respect to the four countries was a relatively low 630 points.

Figure 12 below illustrates the relationship between the original published PASEC proficiency statistics and the adjusted statistics, before the application of grade-related adjustments.

[62] Specifically, the sum of the squares of the differences was minimised.
[63] The mean across five plausible values referring to the reading score, as well the mean across multiple learner weights appliable to each learner, permitted the replication of published PASEC statistics, and was hence the approach used.

Figure 12: PASEC before and after LaNA alignment



Finally, Table 10 displays PASEC-derived proficiency statistics after grade-related adjustments for Gabon and Madagascar were implemented, in line with section 3.3 above.

Table 10: LaNA-aligned PASEC proficiency values

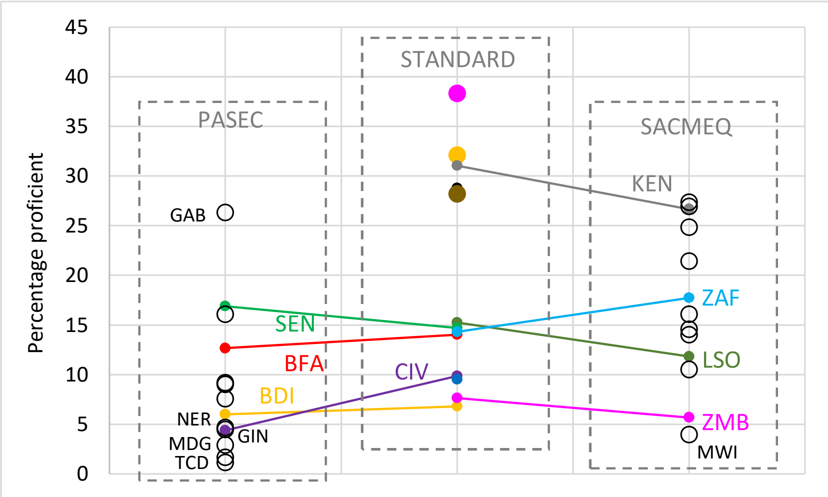
	Reading (≥673)	*Mathematics (≥630)*	*Mean*	*Grade adjustment*
BDI	0.7	11.3	6.0	
BEN	18.4	13.8	16.1	
BFA	7.7	17.6	12.7	
CIV	7.6	1.1	4.4	
CMR	10.6	7.7	9.2	
COD	1.7	1.6	1.6	
COG	10.9	4.2	7.6	
GAB	50.2	22.0	36.1	Yes
GIN	6.1	3.2	4.6	
MDG	2.7	6.3	4.5	Yes
NER	3.5	5.4	4.5	
SEN	13.1	20.7	16.9	
TCD	1.2	1.1	1.1	
TGO	6.1	11.8	9.0	

In the case of PASEC, the across-subject average is simply the average across the reading and mathematics values seen in Table 10. In PASEC the original ranking of reading versus mathematics is retained: most countries have better reading results in both Table 1 and Table 10.

The end result of the PASEC and SACMEQ adjustments, with respect to across-subject averages, is shown in Figure 13. It is noteworthy that even after the adjustments, Gabon emerges as a high performer, not only in terms of PASEC, but also in terms of the broader picture, just below the level of Kenya. The reason why three of the four PASEC countries providing a link to the standard have lower values in the adjusted PASEC box than in the standard box in Figure

13 should be explained. The Burkina Faso standard value is based not just on LaNA, but also AMPL – see the discussion around earlier Figure 9 – and the country performed better in AMPL than LaNA (see Figure 5). The position of Côte d'Ivoire in PASEC is lower than for the standard because this country performs exceptionally poorly in mathematics in PASEC. Of course, for SACMEQ there are also three of four linking countries with adjusted values lower in SACMEQ than in the standard. This comes about largely because of the need, informed by SACMEQ, to maintain a considerable gap between South Africa and Lesotho in the adjusted values, though in the standard the two appear to have almost the same level of proficiency.

Figure 13: PASEC and SACMEQ with adjustments without subject differentiation

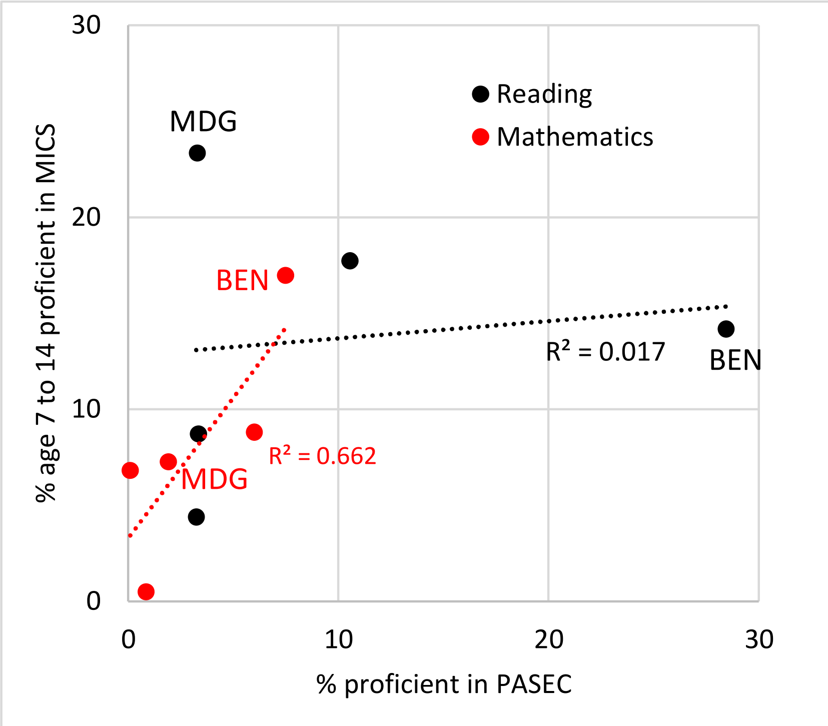


3.6.2 MICS

If the MICS microdata are analysed, it is evident that while published numeracy values are easily replicated and appear reliable, the reading data are not easy to use. Above all, for many countries data are missing for a substantial proportion of children. According to researchers familiar with the data collection process, the fact that a test in an appropriate language had not been produced was a relatively common hurdle. Moreover, consent from an adult in the household to conduct the reading test was fairly often not given. A deeper analysis than that undertaken for the current report is likely to facilitate the imputation of missing reading values using the clearly more available numeracy values[64]. It is worth noting that for the five countries participating in both MICS and PASEC, alignment is fairly high for mathematics and very low for reading. Figure 14 suggests that reading proficiency has been over-estimated in Madagascar's MICS, and under-estimated in Benin's MICS. Different levels of school participation do not appear to be a likely explanation, given that the mathematics values line up relatively well.

[64] A further complexity in the reading data is that to some extent different variables apply, depending on the language of the test. This kind of variation is not found in the numeracy data.

Figure 14: MICS-PASEC alignment in reading and mathematics



Sources: The horizontal axis represents values from Table 10, while the vertical axis represents values from Table 1.

While the above graph suggests strongly that the MICS reading data and statistics should be interpreted with caution, in some ways the MICS reading data display the expected patterns. In particular, the correlation between the MICS reading and numeracy values in Table 1 is .76, essentially the same as the .75 correlation for PASEC, though not has high as the .87 correlation produced by SACMEQ.

With the primary aim of extracting numeracy proficiency values for learners enrolled in grade 6 in the MICS data, microdata were downloaded from the MICS online data repository for 15 of the 18 MICS countries appearing in Table 1. Of the 15, three were selected because they featured among the standard countries of Figure 9 – the three countries were Lesotho, Nigeria and Tunisia. Countries linking MICS to PASEC constituted a set of five additional countries: Benin, Chad, Togo, Madagascar and DRC. Finally, for seven countries MICS was considered indispensable because they had no other data source: São Tomé and Príncipe, Gambia, Comoros, Ghana, Guinea-Bissau, Central African Republic, and Sierra Leone. The three MICS countries whose data were thus not analysed are Malawi, Eswatini and Zimbabwe, all countries with SACMEQ statistics.

Figure 15 illustrates the risk in considering grade 6 learners only up to age 14. While the risk seems negligible in countries such as Tunisia and São Tomé and Príncipe, where over-aged learners are limited, the risk for countries such as Guinea-Bissau and Central African Republic is clear. In these countries, the patterns suggest there are many learners above age 14. Figure 16 illustrates the mathematics performance of grade 6 learners in countries likely to have considerable over-aged enrolment. There is no striking pattern suggesting that learners over age 14 are likely to be much better or worse in mathematics learners up to age 14.

Figure 15: Age distribution of grade 6 learners

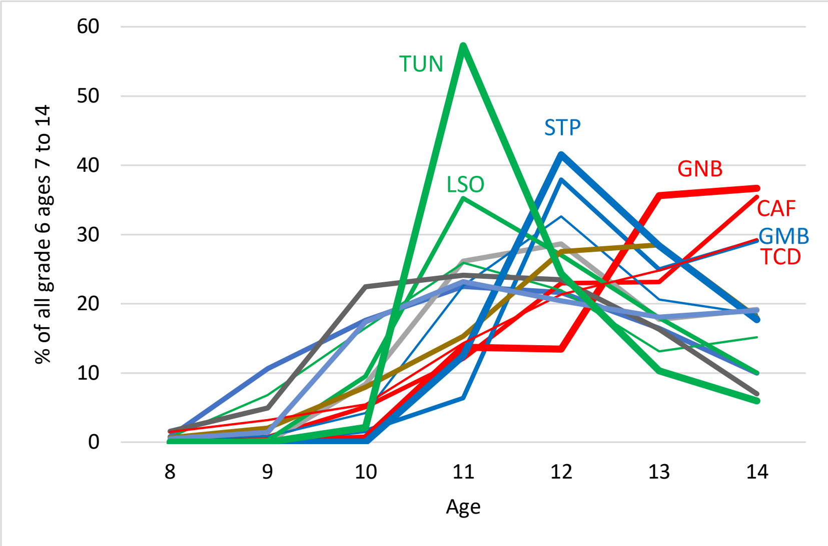
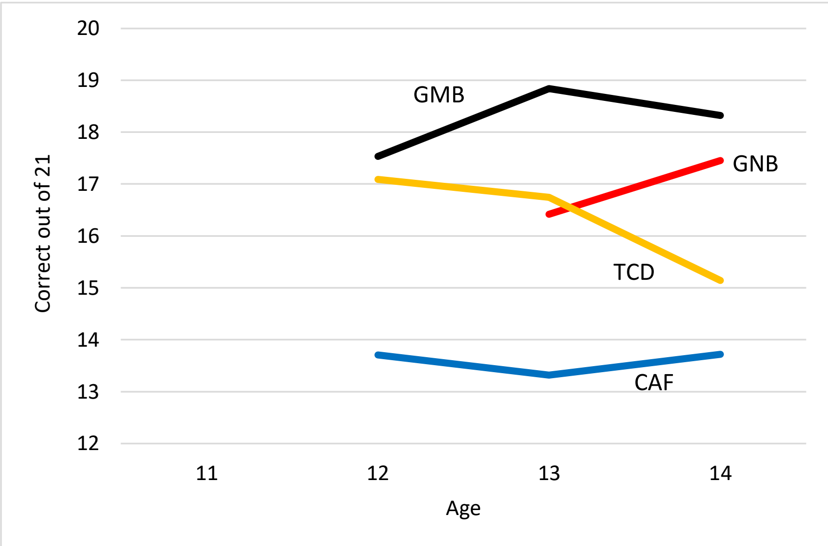


Figure 16: Mathematics performance in grade 6 by age



Note: Age-specific statistics are only shown if there were at least 30 observations behind the statistic. Values in this and the previous graph use the sampling weights in the microdata. Throughout, the basis for identifying grade in the data was a combination of the two variables ED10A and ED10B.

A user manual for the MICS microdata, UNICEF (2020), explains the MICS numeracy assessment data, which are relatively easy to use. The measure of every child's mathematics skills is how many correct responses there were to 21 questions. For a child to be considered proficient for the statistics appearing in Table 1, all 21 questions must be answered correctly. It was possible to replicate the published MICS numeracy statistics of Table 1 with just one minor deviation[65]. Figure 17 compares the Table 1 numeracy proficiency values – on the horizontal axis – to proficiency values for learners up to age 14 enrolled in grade 6 – the vertical axis. Chad (TCD) illustrates the effects of non-participation on the latter statistic. Proficiency

[65] For 14 countries the percentage value was identical, for instance 17% for Benin in Table 1 and 17% based on the microdata analysis. The exception country was Ghana, where 16% seen in Table 1 is slightly higher than the 14% obtained from the microdata.

emerges as high because many children are not in school and those who do attend are often those whose results are relatively good. If all children are counted, Chad is no longer an outlier and the correlation between the two statistics rises considerably – see Figure 18.

Figure 17: Grade 6 mathematics proficiency in 15 MICS countries

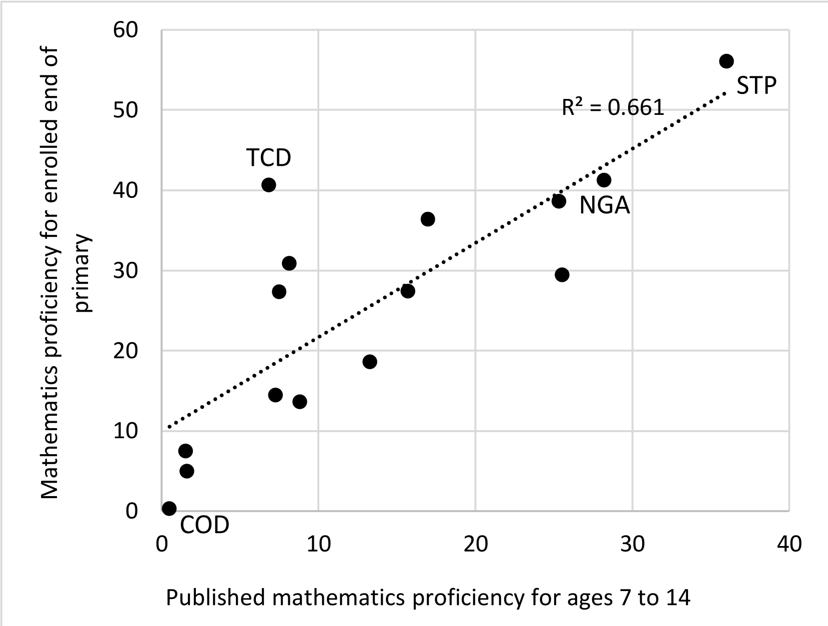


Figure 18: Age 14 mathematics proficiency in 15 MICS countries

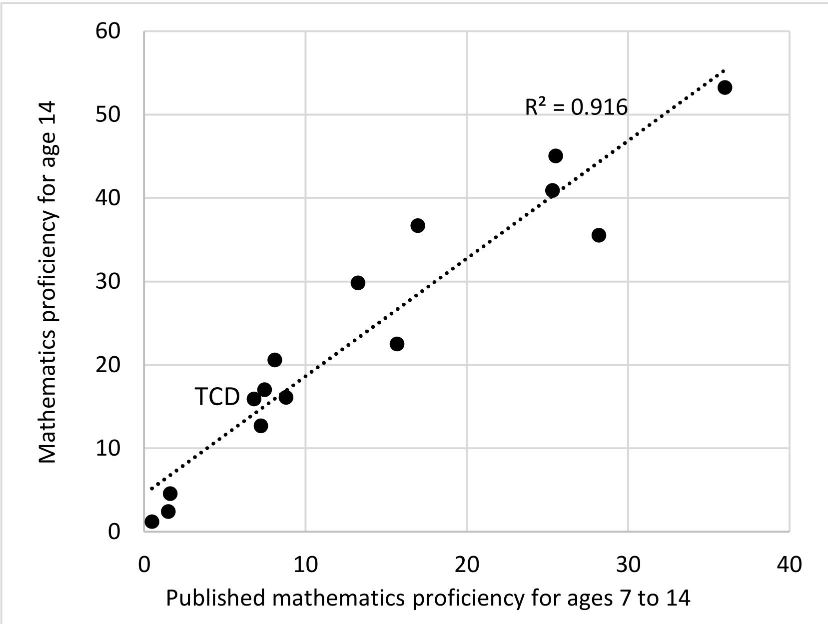
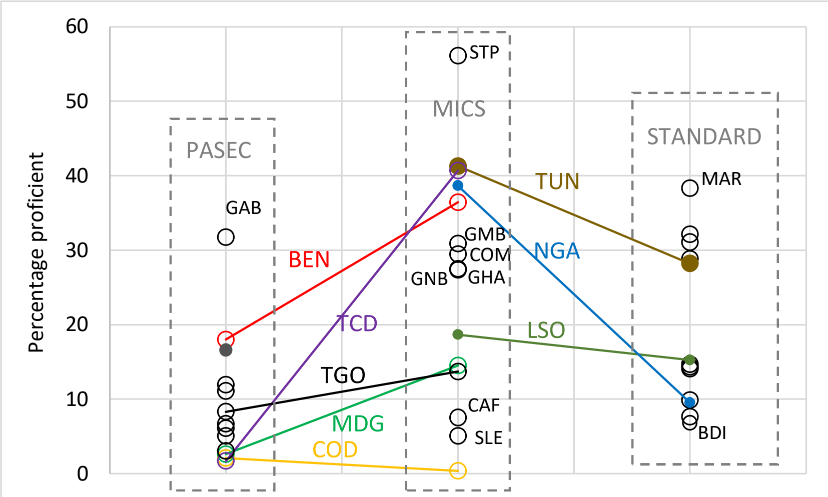


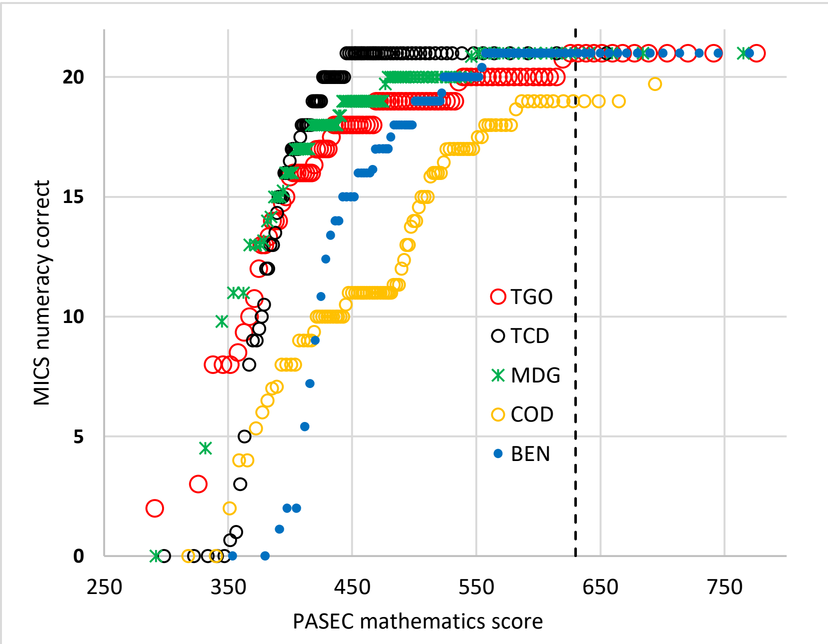
Figure 19 below compares the grade 6 MICS results to PASEC results and the Figure 13 standard values. The approach taken here is to consider MICS numeracy results as the best available MICS-based measure for learning *in general* in MICS countries. Hence the initial focus is on aligning the MICS numeracy values to the across-subject averages of PASEC and of the standard. Given that the MICS assessments are designed for **children** as young as seven, it is not surprising that grade 6 values in Figure 19 should emerge as relatively high.

Figure 19: PASEC-MICS-Standard comparison



The relatively low level of difficulty of the MICS numeracy assessment means that a focus on older children, as is the case here, results in ceiling effects, or little differentiation in the results of better performing learners. This can be seen in Figure 20, where the 630 proficiency threshold for PASEC mathematics identified above (section 3.6.1) corresponds to the maximum score of 21 in MICS for four of the five countries participating in both MICS and PASEC – the exception is DRC (COD). A proficiency threshold of 21 in MICS numeracy would therefore be meaningless: many learners who would obtain less than 630 in PASEC would be considered proficient according to the MICS data.

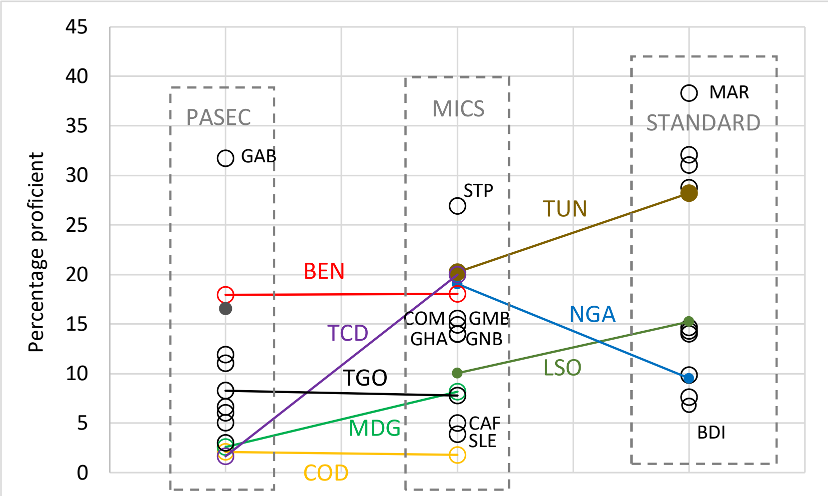
Figure 20: PASEC-MICS concordance in mathematics



Note: The concordance between the two metrics is established by comparing the average score in a percentile in one assessment to the average score in the same percentile in the second assessment. Each marker in the graph thus represents a performance percentile within a country.

Given that there was no microdata solution to adjusting the MICS proficiency statistics, an approach similar to the one employed for the SACMEQ adjustments (section 3.6.1) was used. In other words, a linear transformation of the grade 6 MICS values was performed. The difference is that here alignment against *two* external metrics was pursued: PASEC and the standard values. Each of the two was weighted similarly in the optimisation, as each was considered equally important. Moreover, the optimisation excluded any consideration of Chad, given the very unusual relationship in Chad (TCD) between MICS and PASEC, the proficiency values being 41% and 2% respectively (and with both referring to enrolled grade 6 learners). The adjusted MICS values are shown in Figure 21 below, and the first column of Table 11[66].

Figure 21: PASEC-MICS-Standard comparison after adjustments



Though MICS numeracy microdata were used for the above analysis, these data were used to predict the across-subject averages for the 15 countries. A simple approach was employed to estimate adjusted reading and numeracy scores for MICS. The intention was to retain across-subject gaps as they appeared for MICS in Table 1. To illustrate the method, the across-subject average estimated above for Ghana's grade 6 was 14%. In Table 1 the Ghana reading and mathematics values were 21.4% and 15.7% respectively, giving a mean of 18.5%. The estimated mean of 14% divided by the original mean of 18.5% is 0.75. This implies the Table 1 subject-specific values should also be deflated by a factor of 0.75. This produces the 16% and 12% for the two subjects in Ghana seen in Table 11. More certainty around the relationship between reading and mathematics, specifically among grade 6 learners, could have been achieved had the MICS reading microdata been analysed, but this task was not embarked upon for reasons provided above.

[66] The coefficients applied to the original Figure 19 proficiency values were an intercept of 1.629 and a slope coefficient of 0.451. Optimisation consisted of minimising the sum of a squared differences value in the case of MICS-PASEC and the sum of a squared differences value in the case of MICS-standard. For each of the two values, the mean across the four or three countries was found, thus producing an equal weighting of MICS-PASEC versus MICS-standard.

Table 11: Adjusted MICS values

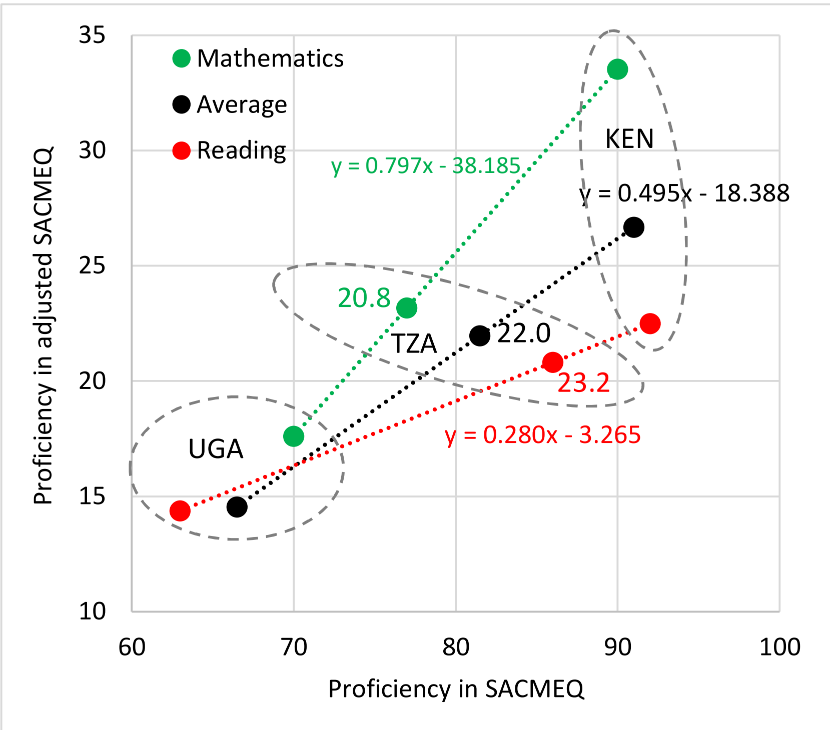
	Average	Numeracy	Reading
BEN	18	16	20
CAF	5	8	2
COD	2	3	0
COM	15	15	15
GHA	14	16	12
GMB	16	19	12
GNB	14	17	11
LSO	10	15	5
MDG	8	12	4
NGA	19	20	19
SLE	4	7	1
STP	27	28	26
TCD	20	16	24
TGO	8	10	5
TUN	20	28	12

Tunisia's 20% average in the above table is notably low, given the country's 28% reflected in the standard (Figure 9). While the 28% should be considered more reliable, the fact that Tunisia's value in MICS is not higher seems to preclude the possibility of major post-2007 improvements in this country (Tunisia's 28% is ultimately based on rather old 2007 data).

3.6.3 PAL

Only one country depends on PAL as a source for adjusted proficiency values, and this is Tanzania. Because in PAL Tanzania is clearly situated between Kenya and Uganda, it was a straightforward process to produce new Tanzania values based on the relationship between new SACMEQ values (from Table 9) and PAL seen in Kenya and Uganda. The approach and new values for Tanzania are illustrated in Figure 22 below.

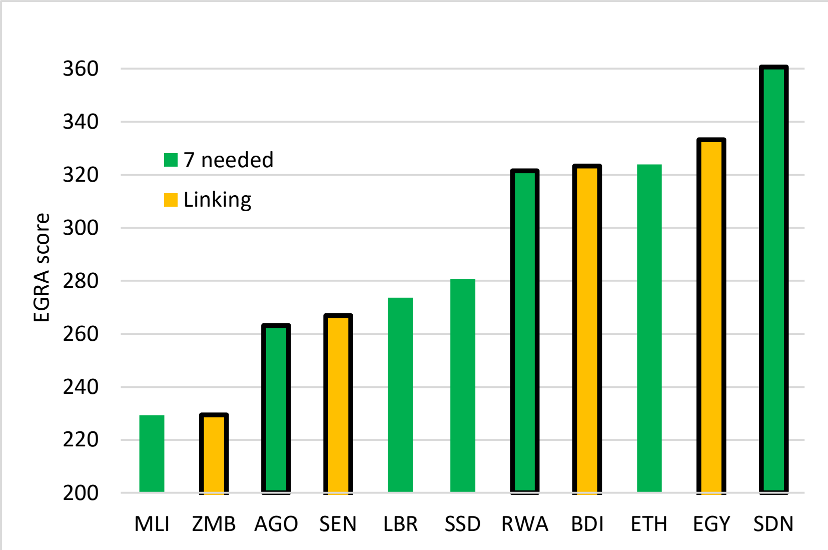
Figure 22: Adjusted proficiency values for Tanzania



3.6.4 EGRA

Figure 23 below draws from Table 1 and Table 4 above. Seven countries for which no alternative to EGRA exists are shown with green bars. Of these, four are reported to have EGRA statistics drawing from a sub-national sample – if there is no solid line around the column, the sample is sub-national. There are also four potential linking countries which can be found among the standard countries of earlier Figure 5 and which are said to have EGRA scores which are nationally representative – see the four yellow bars. Fortunately, the four countries span both the lower and higher end of the performance range.

Figure 23: EGRA mean scores



Key questions include how unrepresentative of the country the four sub-national samples are, and what grade was tested in all of the 11 countries shown in the graph. This information is not available in the spreadsheet accompanying Angrist *et al* (2021) and discussed in section 2 above. Table 12 below is the outcome of an online search where the ideal was to find a report covering specifically the data collection in question for each of the 11 countries of Figure 23. The year of data collection for each country does appear in Angrist *et al* and was reflected in earlier Table 2.

A key finding was that Ethiopia's 2010 EGRA collection does in fact appear to be nationally representative according to the relevant report, though Angrist *et al* decided to mark this collection as not nationally representative. This is fortunate, as Ethiopia accounts for around 46% of children in the '7 needed' countries of Figure 23, and 9% of all Africa's children. This leaves three countries – Mali, Liberia and South Sudan – with sub-national samples. For Mali and Liberia the degree to which the collection was not nationally representative can roughly be known as some details appear in the relevant reports. For South Sudan this could not be ascertained as no report could be found. For Rwanda and Sudan it was possible to confirm through available reports that the data collections were nationally representative. In the case of Angola (AGO), Angrist *et al* indicated that the collection was nationally representative, though this could not be separately confirmed. Of the '7 needed' countries, five have details on the grade tested. For Angola and South Sudan, grade is unknown and hence the need for grade-related adjustments cannot be ascertained. Though Angrist *et al* argued that grade adjustments make little difference to the analysis – see the section 2 discussion above – the evidence on grade-on-grade gains – see section 3.2 – suggests strongly that these adjustments should be undertaken if possible.

Turning to the four linking countries of Figure 23, which are entered in bold text in Table 12, only Zambia (ZMB) appears problematic insofar as the available documentation points to testing have occurred for one language only, or around 20% of the schooling system.

Table 12: Information from EGRA reports

	Report available?	*Nationally representative in Angrist* et al	*Nationally representative in metadata*	*Grade(s)*
MLI 2015	Yes[67]	No	No (±60% of pop., capital Bamako excluded)	2 for some schools, 4 for others (no 3)
ZMB 2011	**Partly**[68]	**Yes**	**No (focus on one lang. ±20% of pop.)**	**2**
AGO 2011	No	Yes	?	?
SEN 2009	**Yes**[69]	**Yes**	**Yes**	**3**
LBR 2013	Partly[70]	No	No (4 of 15 counties covered)	1 to 3
SSD 2017	No[71]	No	?	?
RWA 2016	Yes[72]	Yes	Yes	1 to 4
BDI 2012	**Yes**[73]	**Yes**	**Yes**	**2 to 3**
ETH 2010	Yes[74]	No	Yes	2 to 3
EGY 2013	**Partly**[75]	**Yes**	**Yes**	**3**
SDN 2015	Partly[76]	Yes	Yes	3

The last column of Table 12 suggests grade-related adjustments should occur in the case of Zambia, Liberia, Burundi and Ethiopia. As discussed in section 2, Angrist *et al* ignored data from outside the grades 2 to 4 range, so the question is whether the selection of grades within that range is likely to inflate or deflate proficiency statistics. Rwanda would not be distorted as the full grades 2 to 4 range would have been used. Countries such as Egypt with only grade 3 would also not be distorted, assuming that grades 2 and 4 cancel each other out. In the case of Zambia and Liberia the average grade covered is grade 2, when the norm is implicitly grade 3. This suggests the mean values for these two countries should be inflated by one grade's worth of learning, which section 3.2 above suggests should be around 30 score points. This estimate of an adjustment is based on the knowledge that the standard deviation in Angrist *et al*'s EGRA means is 100 (see section 2 above). Following the same logic, the average grade tested in Burundi and Ethiopia was 2.5, necessitating a 15-point upward adjustment for these two countries.

Using only the four linking countries referred to above to convert EGRA means of the seven new countries to proficiency statistics along the metric of the standard resulted in new

[67] RTI International, 2015.
[68] While no report could be found for the 2011 testing, a report on a later 2014 wave of testing in Zambia, which was nationally representative, refers briefly back to the 2011 data. See Brombacher *et al* (2015).
[69] Pouezevara *et al*, 2009.
[70] While no report could be found for the 2013 testing, a report on a later 2015 wave of testing in Liberia provides details on the earlier 2013 wave. See King *et al* (2015).
[71] While some reports on South Sudan EGRA testing occurring before 2017 could be found, almost nothing could be found for 2017 testing. A significant part of the problem seemed to be that the website for the consulting firm Montrose International was no longer functioning. Cambridge Education (2019: 53) seemed to confirm that testing had occurred in 2017, but without providing details.
[72] Education Development Center, 2017.
[73] Mauzunya and Varly, 2012.
[74] Piper, 2010.
[75] While the report for the 2013 assessments could not be found, RTI International (2014) deals with the 2014 assessments and refers back to the 2013 data collection.
[76] World Bank (2019: 53) makes reference to the 2015 data collection.

proficiency values which were counter-intuitively high. Among the problems of this approach was that Burundi was ranked very differently in the standard compared to the EGRA countries of Angrist *et al*. In Angrist *et al* Burundi was ranked high, almost at the level of Egypt and well above that of Senegal – see Figure 23. In the standard, Burundi has been ranked well below Senegal – see Figure 9. Moreover, in Angrist *et al* the mean for Sudan is considerably above that for Egypt (see Table 1), a counter-intuitive pattern given that Egypt's per capita income has been around twice that of Sudan in recent years. Given this situation, it was decided to examine the relationship between all Angrist *et al* means and harmonised proficiency values derived so far in section 3.4 above and to use that relationship to convert EGRA to proficiency values in the case of the seven countries. Given the apparent utility of across-subject average values, it was furthermore decided to use EGRA to predict only the across-subject average proficiency values. Fortuitously, the relationship between Egypt and Sudan was reversed in the process, with Egypt's proficiency exceeding that of Sudan by a considerable margin in the new rankings.

Figure 24 illustrates the relationship. The equation is based on a quadratic function covering 34 countries available in Angrist *et al*, but also in one of the large assessment programmes discussed above (all except for PAL). A conversion of means to proficiency scores can be expected to use a quadratic function. For the horizontal axis, the most recent reading mean was found. For the vertical axis, some countries had more than one harmonised value in the work presented above. In such cases, the value in the standard was preferred, followed by values in PASEC, SACMEQ and MICS. Moreover, the seven new countries, with predicted proficiency values based on the function, are shown using red markers, with their values reported in Table 13.

Figure 24: Angrist* et al *means and harmonised proficiency values

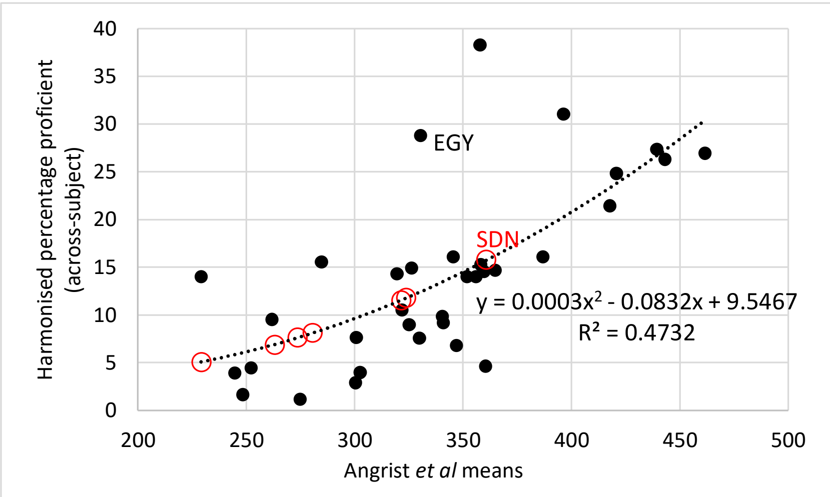


Table 13: Harmonised EGRA-based proficiency values

Country	Proficiency (across-subject average)
AGO	6.9
ETH	11.8
LBR	7.6
MLI	5.1
RWA	11.5
SDN	15.8
SSD	8.1

It should be noted that the 2023 national assessment of Rwanda[77] referred to in section 2 produced proficiency values which were much higher than what is seen in Table 13. Most notably, 91% of grade 6 learners were found to be proficient in mathematics, despite the use of non-statistical 'policy linking' to align the assessment to UNESCO standards. The value of 91% would not only place Rwanda well above any other African country. It would also put Rwanda's proficiency well above that of developed countries such as Finland[78].

3.7 Step 5: Selection of best statistics per country and a reality check

Table 14 below draws from the above analysis in presenting final statistics. The choice of best statistics per country is straightforward and mostly follows the logic of the previous sections, where the approach was to identify the standard, and then align values from other programmes, while moving from more to less reliable data sources.

There were thirteen countries with statistics available from sources considered to follow the SDG standard most accurately. The key graphs above in this regard are Figure 6 and Figure 9. There were three countries for which Figure 13 indicates that following the standard would lower the statistics substantially, and worsen the ranking, relative to what is seen in PASEC or SACMEQ. The three countries are Senegal, Burundi and South Africa. For these, it was decided to use adjusted PASEC or SACMEQ values. This explains the two instances in PASEC and one in SACMEQ with an asterisk in Table 14. It also explains why Table 15, which summarises the data sources, refers to just ten countries drawing from the standard.

The PASEC and SACMEQ values in Table 14 draw from Table 9 and Table 10 above. MICS serves as a source for seven countries in Table 14, with adjusted MICS values being from earlier Table 11. PAL was needed for just one country, Tanzania – see earlier Figure 22. Finally, seven countries drew from EGRA, where reading figures were used to generate average proficiency values – see Table 13 above. Given the finding that in many respects averages across reading and mathematics are more reliable than values for the two subjects viewed separately, the decision was taken to use EGRA to augment the set of averages, rather than to inform just reading values.

[77] Rwanda: National Examination and School Inspection Authority, 2023.

[78] In 2023 TIMSS, 76% of Finland's grade 4 learners exceeded the intermediate benchmark, which UNESCO has indicated is the relevant threshold to use.

Table 14: Final harmonised proficiency statistics

Country	*Source*	*Average*	*Reading*	*Math.*
BFA	Standard	14.0	12.4	15.7
CIV	Standard	9.9	10.8	8.9
DZA	Standard	32.1	50.2	14.0
EGY	Standard	28.8	37.9	19.7
KEN	Standard	31.1	25.2	36.9
LSO	Standard	15.3	10.8	19.7
MAR	Standard	38.3	54.6	22.0
NGA	Standard	9.5	12.7	6.3
TUN	Standard	28.2	47.4	9.0
ZMB	Standard	7.6	5.5	9.8
BDI	PASEC*	6.0	0.7	11.3
BEN	PASEC	16.1	18.4	13.8
CMR	PASEC	9.2	10.6	7.7
COD	PASEC	1.6	1.7	1.6
COG	PASEC	7.6	10.9	4.2
GAB	PASEC	36.1	50.2	22.0
GIN	PASEC	4.6	6.1	3.2
MDG	PASEC	4.5	2.7	6.3
NER	PASEC	4.5	3.5	5.4
SEN	PASEC*	16.9	13.1	20.7
TCD	PASEC	1.1	1.2	1.1
TGO	PASEC	9.0	6.1	11.8
BWA	SACMEQ	21.4	19.7	25.4
MOZ	SACMEQ	10.5	10.1	14.9
MUS	SACMEQ	27.4	21.8	36.3
MWI	SACMEQ	4.0	3.8	9.4
NAM	SACMEQ	16.1	17.6	16.0
SWZ	SACMEQ	24.8	24.5	25.8
SYC	SACMEQ	26.9	23.1	33.1
UGA	SACMEQ	14.5	14.4	17.6
ZAF	SACMEQ*	17.7	16.4	22.0
ZWE	SACMEQ	14.0	12.8	18.9
CAF	MICS	5.0	7.6	2.5
COM	MICS	14.9	14.9	14.9
GHA	MICS	14.0	16.2	11.9
GMB	MICS	15.6	18.8	12.3
GNB	MICS	14.0	17.4	10.5
SLE	MICS	3.9	7.1	0.7
STP	MICS	26.9	27.8	26.1
TZA	PAL	22.0	20.8	23.2
AGO	EGRA	6.9		
ETH	EGRA	11.8		
LBR	EGRA	7.6		
MLI	EGRA	5.1		
RWA	EGRA	11.5		
SDN	EGRA	15.8		
SSD	EGRA	8.1		

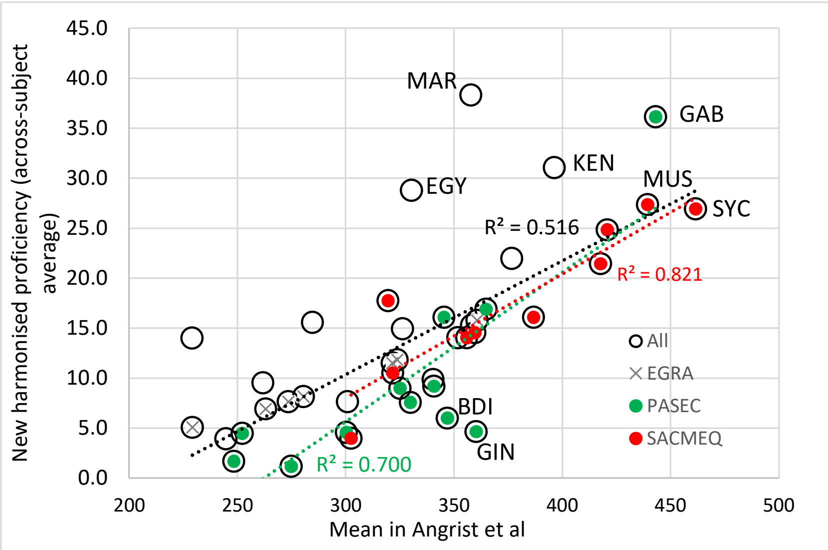
Table 15: Breakdown of countries per source

Approach or source	*Countries*
Standard	10
PASEC	12
SACMEQ	10
MICS	7
PAL	1
EGRA	7
Total countries	47

The figures from Table 14, when combined with child population figures, allow for an Africa-wide aggregate proficiency statistic to be calculated. The across-subject average for the continent is 13.4%, considering countries with 97% of the continent's children. The subject-specific proficiency values are 16.1% for reading and 12.5% for mathematics, though here only 79% of population-weighted Africa is considered.

The 47 countries referred to in Table 15 account for 97% of Africa's children. As discussed in section 3.1 above, Angrist *et al* (2021) also present harmonised values, though they are mean scores and not proficiency statistics, for 94% of the continent. What does a comparison of the Table 14 statistics against Angrist *et al* reveal? Figure 25 represents 42 countries found in Angrist *et al* with a primary-level reading score and also with an average proficiency statistic in Table 14 above. The Angrist *et al* statistic for the most recent year was used where values for more than one year existed. Though there are 45 African countries in the Angrist *et al* spreadsheet, of these 43 have reading scores and just 17 have mathematics scores[79]. The overall correlation for the 42 countries is not high – R squared is .516. However, counting only countries drawing from SACMEQ data in Table 14 yields a fairly high correlation, with R squared being .821. Counting just PASEC countries produces a somewhat lower R squared of .700. Outlier countries in Figure 25 receive attention below.

***Figure 25: Comparing new harmonisation to Angrist* et al**



The fact that the across-subject average of Table 14 is a relatively robust measure is confirmed by the fact that Angrist *et al*'s reading means are more correlated to the Table 14 average than to the Table 14 reading score. Specifically, R squared values of .484 and .271 emerge, using in each instance the same 35 countries (the seven EGRA countries from Table 14 are excluded).

Figure 26 below represents the average values from Table 14 in a schematic map format. The five categories roughly produce five equal counts of countries, meaning the top category is roughly the top quintile (without population weighting).

[79] If only countries with both reading and mathematics values in Angrist *et al* are considered, then the coverage of Africa drops from the aforementioned 97% to just 24%.

Figure 26: Average proficiency per country

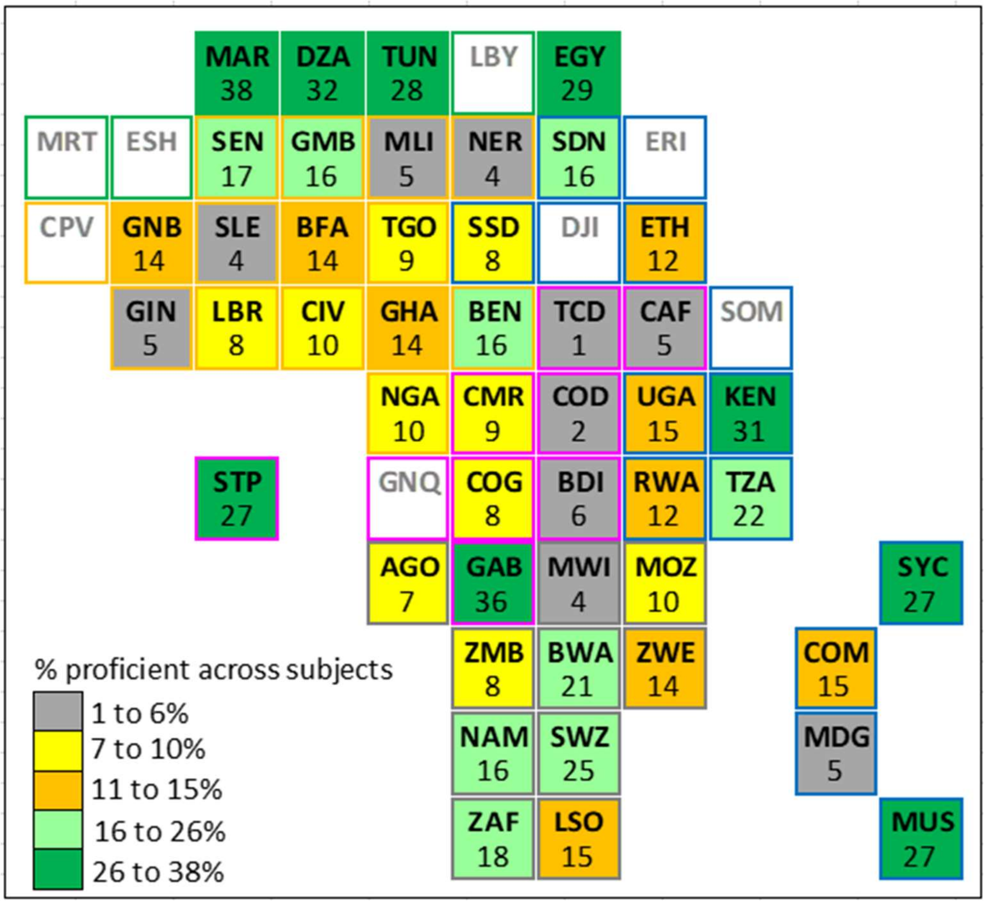
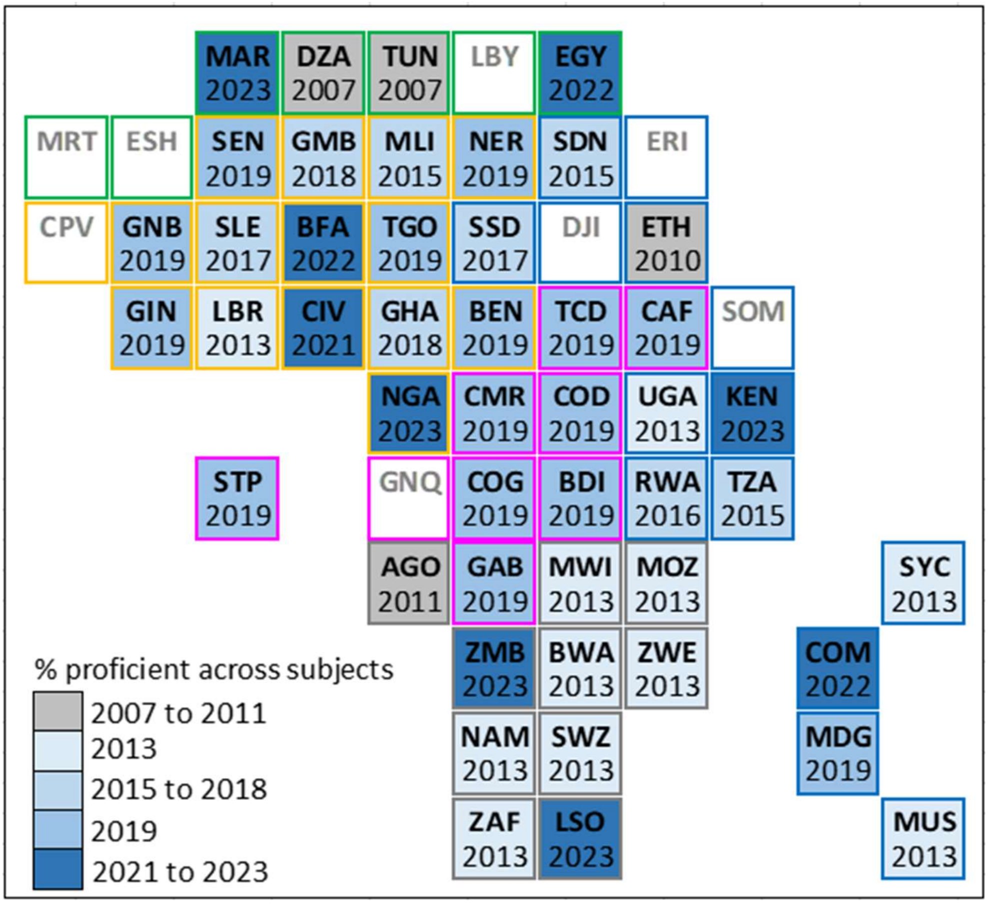


Figure 27 below illustrates the source year for the values of Table 14. In general, the source year should illustrate the risk of recent improvements not having been taken into account. However, this is not always true. South Africa has clearly improved (see section 6), yet small inconsistencies in measurement approaches meant the older SACMEQ harmonised value was preferred to the harmonised (and grade-adjusted) value derived from more recent TIMSS and PIRLS data. This was discussed above.

Figure 27: Source year per country



According to Figure 27, 26 of 47 countries with a data source have a source that is from 2018 or later. This thus represents an substantive update relative to Angrist *et al*, whose most recent reading data source for Africa was 2017.

The following two diagrams illustrate the sources for the final proficiency statistics. Figure 28 focusses on the standard initially discussed in section 3.5. Countries in bold are linking countries spanning more than one source programme. An asterisk means that a grade-related adjustment took place in a context where different countries in the same programme were testing different grades.

Figure 28: Determination of the standard

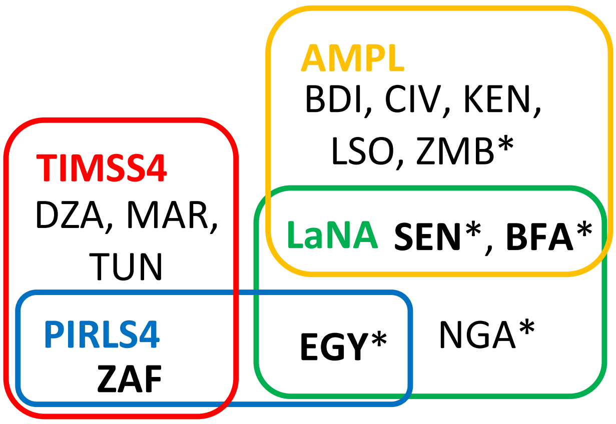
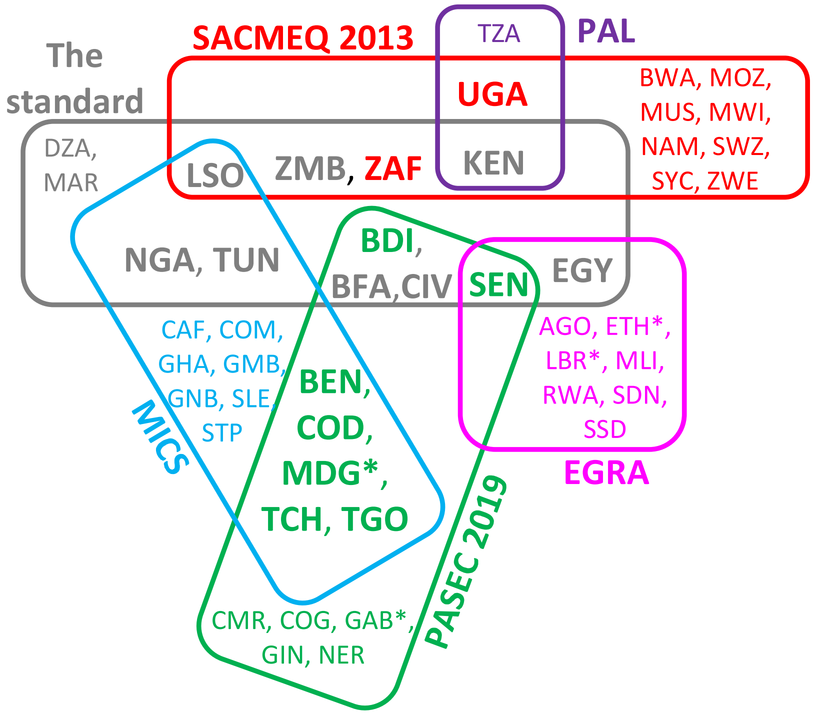


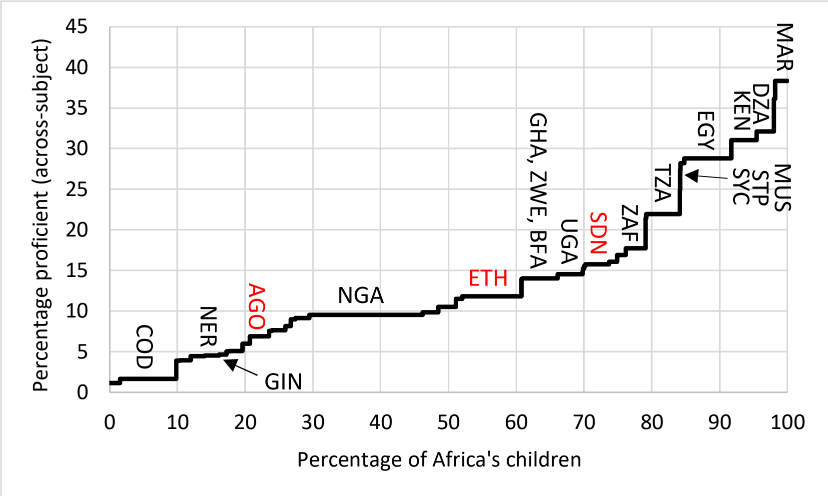
Figure 29 illustrates how other programmes, beyond the standard, were brought into the harmonisation. A larger font for a country means the country served as a linking country. The colours applied to the countries indicate the source for the final proficiency statistics. Three countries in the standard, namely South Africa, Burundi and Senegal, used statistics from programmes outside the standard, for reasons provided above. For this reason, those three countries are marked in the colour of the programme finally used.

Figure 29: Sources for all 47 countries



The harmonised proficiency statistics of Table 14 were presented to a few specialists, who found the values and rankings on the whole plausible, though questions were raised about some countries. The discussion that follows addresses these questions, while using Figure 30 to assist the interpretation. This graph both ranks key countries, and weights all countries by their child population.

Figure 30: Average proficiency across population-weighted countries



Note: Red means EGRA is the source used. As discussed elsewhere in the report, for 3% of the continent's children there were no assessment data to draw from. Hence 100% in this graph actually represents 97% of children.

Egypt ranks far better in the harmonised statistics presented here than in Angrist *et al* (2021) – see Figure 25 above. This is largely due to the fact that Egypt's PIRLS 2016 results used by Angrist *et al* seem strikingly low. The percentage of learners reaching the low international benchmark in PIRLS moved from 31% in 2016 to 45% in 2021. Even without the COVID-19 pandemic, this would be a remarkable and virtually unprecedented speed of improvement, with the PIRLS *mean* rising by 10 points a year. The final 37.9% value for reading in Table 14 in the case of Egypt is the result of averaging across PIRLS 2021 and LaNA – having two separate sources here improves the certainty around Egypt's final figures. Egypt's ranking in the ninth decile in Figure 30 is arguably not surprising, and is roughly in line with Gust *et al* (see section 4).

Morocco ranks far better in the values presented above than in Angrist *et al* (see MAR in Figure 25). As with Egypt, this can be explained by improvements in PIRLS between 2016 and 2021, though these improvements were more plausible in the case of Morocco, at around a gain of 3 points per year in the mean. As pointed in out section 6, Morocco has been highlighted as an exceptional improver in Africa.

Kenya's relatively good position in Figure 30 is unlikely to surprise those accustomed to seeing Kenya near the top of the SACMEQ rankings (see earlier Figure 10). However, Mauritius and Seychelles display slightly better averages than Kenya in SACMEQ (see Table 9), while in Figure 26 these two island nations are placed slightly below Kenya. This is because Kenya is given a slight 'push' above its SACMEQ values by AMPL, which is a more recent source than SACMEQ. It is possible improvements in Kenya come through which cannot be gauged for SACMEQ countries, given the SACMEQ values used are from 2013.

São Tomé and Príncipe's relatively good position in Figure 30, between Seychelles and Mauritius, may be surprising, in part because this country has apparently never featured in rankings of learning outcomes due to a lack of data. The country's inclusion here is possible due to MICS. Even Gust *et al* exclude this country when imputing values from non-assessment data. This country's ranking is not surprising if one considers that the Human Development Index of the UNDP, which excludes measures of actual learning, has in recent years given it a value slightly above that of Kenya. São Tomé and Príncipe moreover ranks highly with respect to indicators of democracy.

South Africa's ranking in Figure 30 in only the eighth decile may be surprising if one takes into account the country's relatively good performance in university rankings[80]. However, South Africa's poor averages, driven by high levels of inequality, are well-documented and supported by multiple sources of evidence. Yet South Africa, like Morocco, has been considered a fast improver in the area of learning outcomes, albeit off a low base.

Guinea emerges with a better ranking in Angrist *et al* than in the new values – see GIN in Figure 25. This would be related to the fact that Guinea is one of only a few countries where PASEC 2006 results were used in Angrist *et al*. For PASEC, Angrist *et al* mostly used 2014 data – the 2019 PASEC results available for the current paper were clearly not available then. It appears that PASEC 2006 values are seldom if ever considered comparable to 2014 or 2019 PASEC values, even within PASEC's own reports.

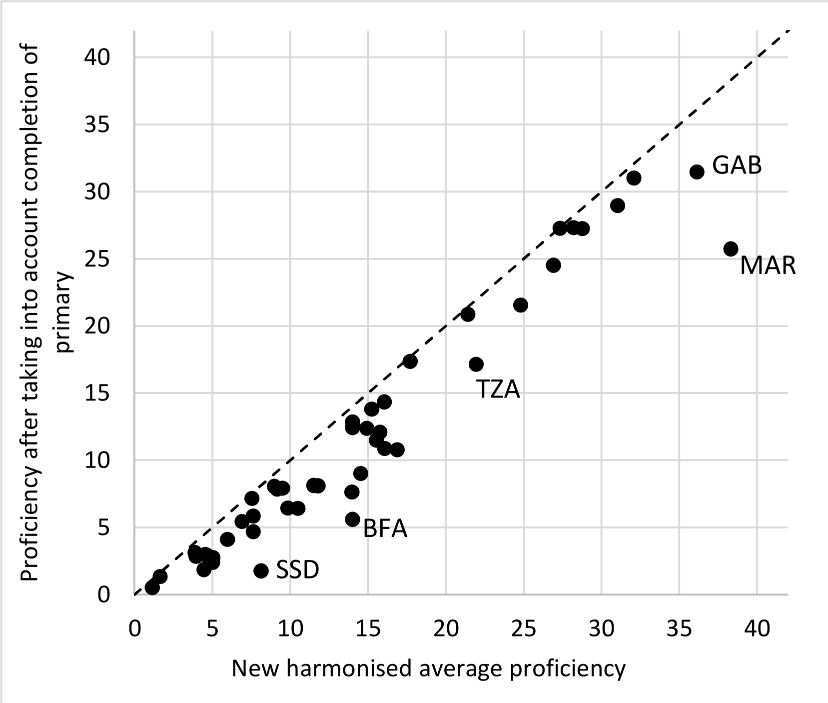
4 How considering the out-of-school changes the picture

The definitions of the SDG 4.1.1 indicators, and actual SDG-related statistics, exclude children not in school. Yet there is inevitably much interest in estimating proficiency levels in the young population as whole, regardless of school participation. Population is thus the ultimate focus of Gust *et al* (2024). UNESCO has recommended using the reasonable assumption that, for instance, children who should be in grade 6 and who are not do not reach minimum proficiency levels associated with this grade[81]. Such an approach is more complex than may seem to be the case if one considers that in many countries both participation and learning are substantially delayed. Thus, in the case of a child aged 13 who should be in grade 6 but is still in grade 3, the possibility cannot be excluded that the child will reach grade 6 and reach the associated minimum proficiency level. Moreover, a child in grade 6 who has not reached the minimum level associated with the end of primary may do after some years later, at the secondary level.

Figure 31 partly deals with the delay issues by only counting someone as not reaching the end of primary if he or she did not reach this level eight years after the ideal age. The data source for this is Dharamshi *et al* (2021), who provide a particularly comprehensive standardised dataset of school completion drawing from household data. The most recent year per country for Africa ranges from 2005 to 2020, though two-thirds of countries have values for 2020. A country where all children complete primary schooling would lie on the diagonal. Countries to the right of the diagonal are those where a significant number of out-of-school children exist, using the eight-year criterion, children we assume do not attain the end of primary proficiency level. This would lower the across-subject average proficiency values reported in earlier Table 14.

[80] The World University Rankings of 2025 place four South African universities among the top 500 in the world, and one from Morocco. Only these two countries from Africa make it into this top 500 (www.timeshighereducation.com).
[81] UNESCO Institute for Statistics, 2017.

Figure 31: New proficiency estimates and out-of-school adjustments



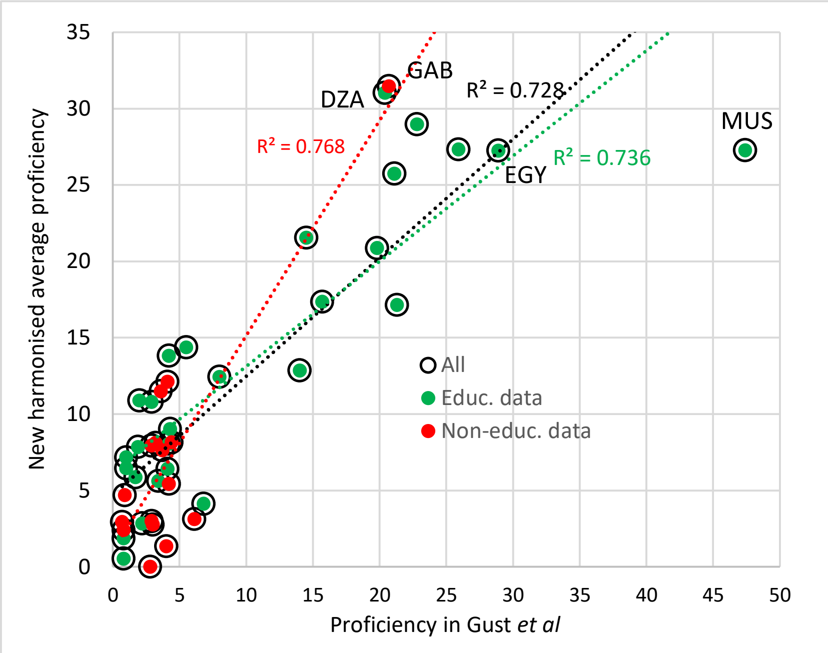
Note: There are 46 countries represented in the graph. Seychelles is missing in the Dharamshi et al *(2021) data.*

The proficiency values represented along the horizontal axis of Figure 31 produce an aggregate of 13.4% when child populations are used as weights. This drops to 10.8% if out-of-school adjustments are taken into account. Put differently, only 11% of young Africans attain the minimum proficiency level for end of primary as understood within the SDG monitoring system.

The fact that Morocco displays relatively good proficiency levels in the schooling system, yet low levels of participation, is striking. The Dharamshi *et al* dataset points to only 67% of young Moroccans not reaching the end of primary schooling. However, this is based on 2009 data, making Morocco's statistic among the most dated in the dataset. World Development Indicator values for the 'Primary completion rate' place Morocco in a favourable position in Africa in recent years, on a par with South Africa and Tunisia. In short, Morocco's placement in Figure 31 is probably not reliable.

Figure 32 compares the adjusted values from Figure 31 to proficiency values in Gust *et al* (2024). There are 44 countries common to both. R squared considering all countries is .728, which rises only slightly to .736 if only the 28 countries in Gust *et al* drawing from education data are considered. These values are reassuringly high, especially if one considers that Gust *et al* draw mostly from older data, in particular earlier PASEC and SACMEQ. The correlation is not low for countries where Gust *et al* used non-assessment data – R squared is .728. This suggests that non-education data are relatively good at predicting learning outcomes, even if ideally data on learning outcomes should be used when analysing learning outcomes.

Figure 32: Comparison of new estimates to Gust et al

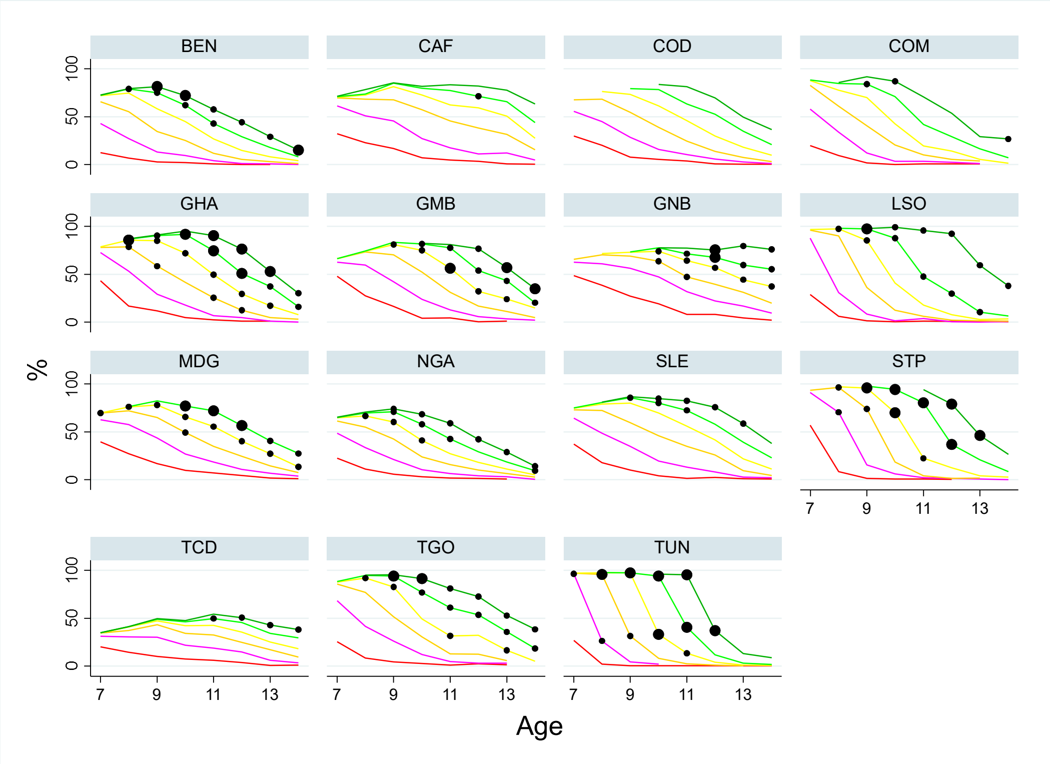


5 How lower primary predicts end of primary

The current report has focussed on the harmonisation of proficiency statistics at the end of the primary level. What do these statistics say about learning occurring in the earliest grades, for instance as measured by SDG indicator 4.1.1a, whose focus is grades 2 and 3?

Answering this question well requires having a sufficient picture of how children are distributed across grades. How good is school participation at specific ages, and what does the distribution of children across the school grades look like? The latter depends to a large degree on the extent to which children enter grade 1 early or late, and the extent of grade repetition. Figure 26 below illustrates how irregular the distribution across grades is in many African countries. The graph uses MICS data for the 15 countries appearing in earlier Table 11. The pattern in Tunisia is what would often be considered the ideal. There is little grade 1 enrolment above age 7 and by age 13 all the enrolled are beyond grade 5. In many countries there are high numbers of children aged well above age 7 in grade 1 – for instance in Guinea Bissau (GNB) 69% of grade 1 learners are older than 7, against just 9% in Tunisia. Moreover, by age 14 a high proportion of learners are still in the first three grades – this is true for a quarter of age 14 learners in Madagascar.

Figure 33: Grade-participation-numeracy relationships



Source: Analysis of MICS microdata.
Note: The six curves represent the six grades in the 1 to 6 range. They show cumulative enrolment as a percentage of the population. Thus the top and dark green curve shows enrolment in all grades up to grade 6. The bottom red curve shows just grade 1 participation. Madagascar has only five primary grades, and thus displays only five curves. The black markers represent levels of numeracy. There are two sizes. The smaller markers indicate that a mean of at least 15 of 21 correct is achieved for that age and grade. The larger markers use a threshold of 18 correct. Markers are only shown if there are at least 30 observations (or tested children) for the specific country, age and grade. For 24% of the country-age-grade groups there were fewer than 30 observations.

Figure 34 below, which also draws from MICS, points to two important things about the relationship between learning outcomes in the initial grades and learning in grade 6. Firstly, if children display relatively good skills already in grades 1 or 2, the country is likely to display relatively good skills in grade 6. Put differently, some countries have a more favourable point of departure. To illustrate, while on average children in Comoros (COM) gain skills as they move up the grades, these gains are much smaller than the gap between the highest and lowest proficiency values across the countries at the grade 1 entry point. The second thing that can be seen is that in some countries the gains across grades are greater than in other countries – the lines cross each other. This could be because the schooling system is effective, though it could also be because academically weaker children drop out of school[82].

[82] Silberstein (2021), in one of the few published cross-country analyses using MICS foundational learning data, takes both participation and grade into account by examining skills relative to the highest grade attained at any point in the past. The current analysis focusses on the grade in which the child is currently enrolled.

Figure 34: Gains by grade

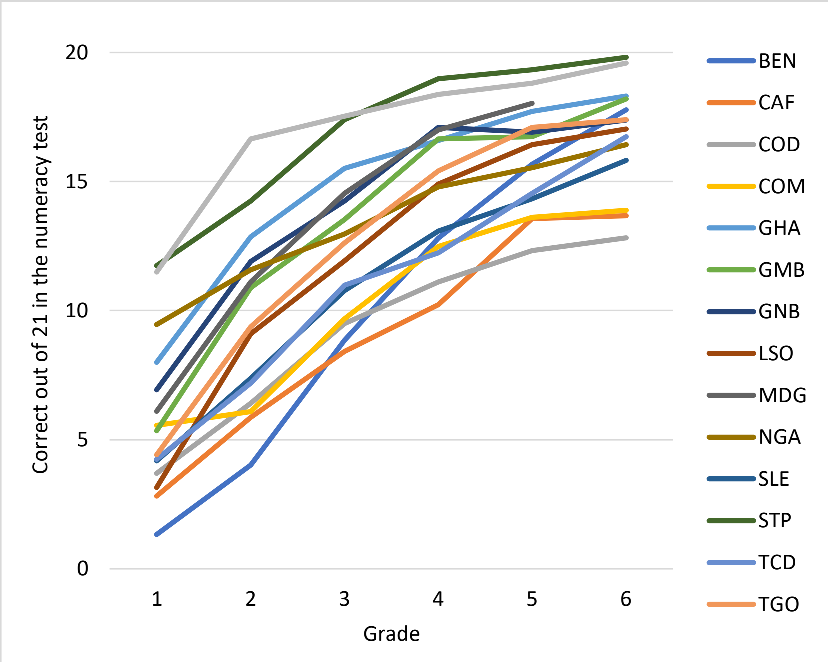
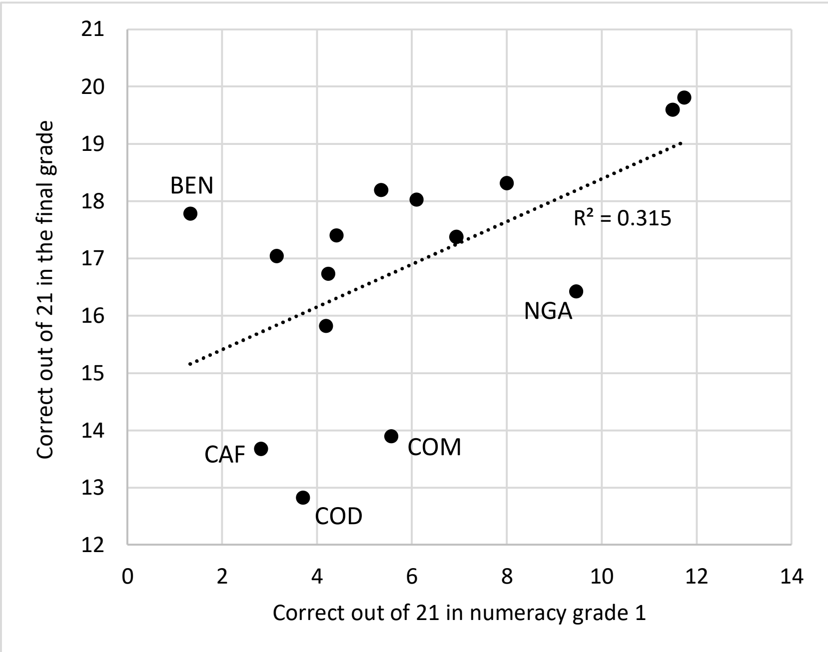


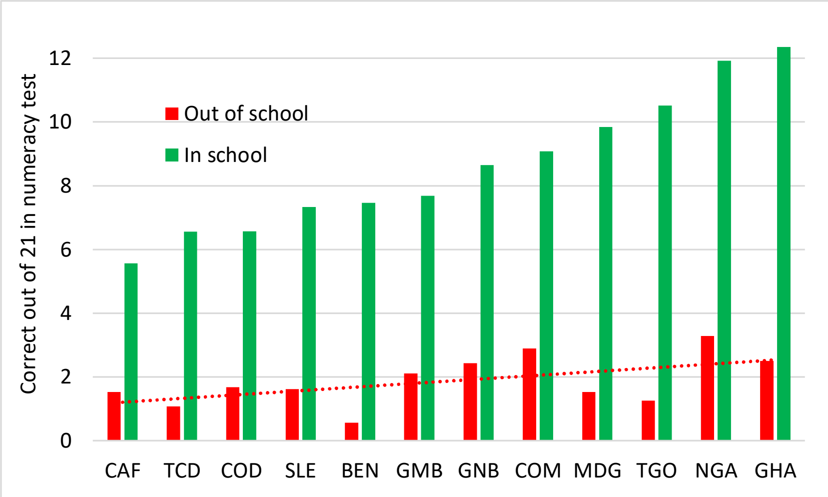
Figure 35 compares grades 1 and 6 (or 5, in the case of Madagascar), repeating values shown in earlier Figure 34. The positive outlier Benin displays above average gains, while Nigeria, Comoros, DRC (COD) and Central African Republic display below average gains.

Figure 35: Gain between grade 1 and grade 6



To what extent are the differences in grade 1 across countries seen for instance in Figure 34 due to children entering school with better capabilities because of factors such as nutrition and pre-school educational activities, at home or in a pre-school? Figure 36, which reflects countries with at least 30 observations of children aged 7 to 8 who are *not* in school, should be interpreted cautiously[83], yet it suggests that much of the proficiency difference across countries among enrolled learners in the early grades is due to what is learnt right at the start of schooling. There is a positive correlation between the green and red bars, or between the skills of children in school and out of school[84], yet differences across the red bars are dwarfed by differences across the green bars.

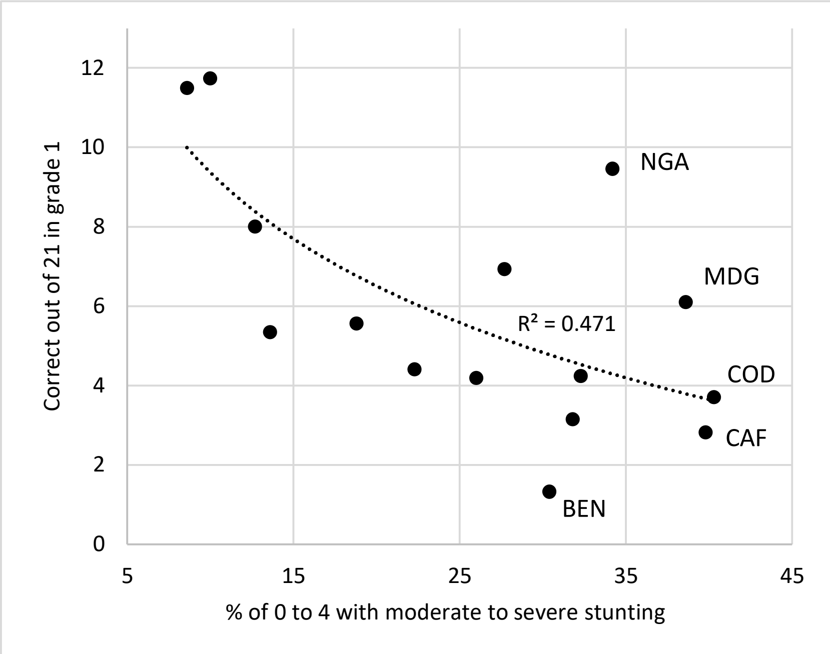
Figure 36: Numeracy skills among those aged 7 and 8



Yet non-educational factors would explain at least some of the grade 1 performance differences. Evidence from Ethiopia points to children at age 8 suffering from stunting achieving 16% correct in a mathematics test, against 25% for non-stunted children, after controlling for various background factors[85]. Figure 37 below provides a reminder that rates of stunting among young children remain high in many African countries. The correlation with learning outcomes is not necessarily a reflection of the size of the impact of stunting on learning. Countries with high levels of stunting may also be countries where teaching and learning in primary schools is relatively weak.

[83] The gap between the red and green columns could be the result of factors other than schooling, for instance disability.
[84] Pearson correlation of .56.
[85] Woldehanna *et al*, 2017.

Figure 37: Stunting and grade 1 numeracy results



Source: For the horizontal axis, values are taken from UNICEF (2024). For the vertical axis, grade 1 values in Figure 34 were used.
Note: The trendline is logarithmic.

An effective way of gauging how well a country's schooling system performs overall, taking both participation and the quality of schooling into account, is to examine learning gains over age, regardless of school participation. This is presented in Figure 38. Across all countries, children become more numerate as they age. The fact that the curves are steepest for younger children is in part a result of ceiling effects, or the easiness of the test, resulting in poor differentiation among older children.

Figure 38: Numeracy skills by age

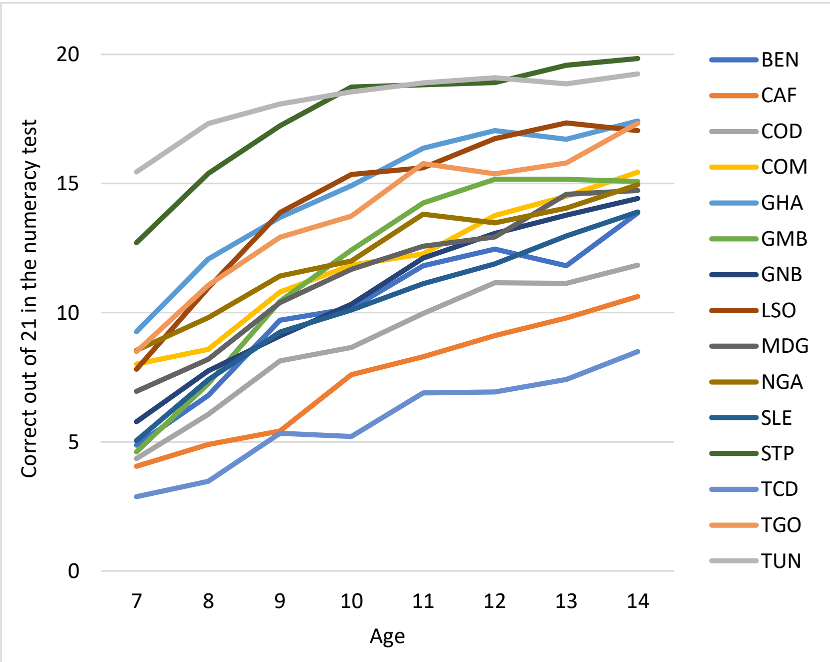
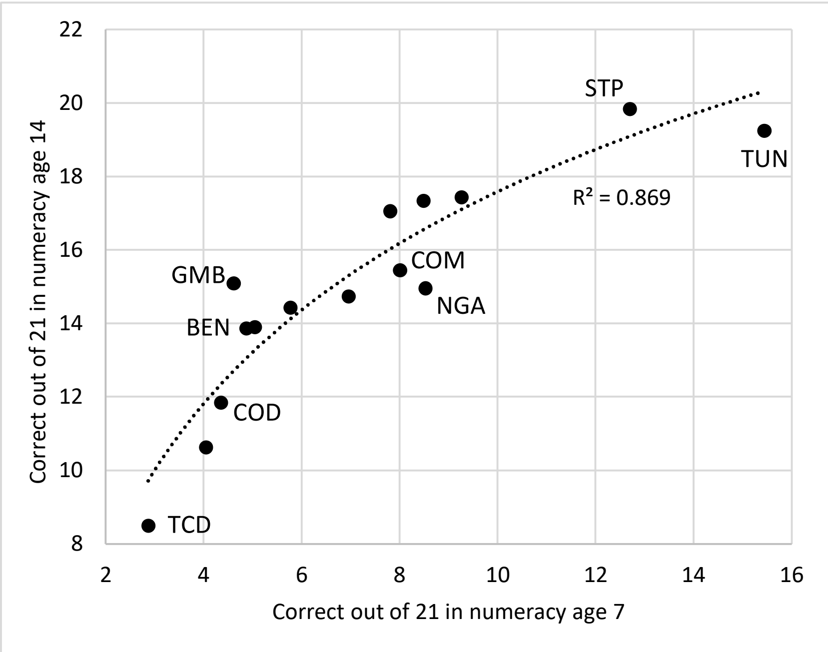


Figure 39 below reflects only ages 7 and 14 statistics from the previous graph. What is striking is how strongly skills already present at age 7 predict skills seen at age 14. Despite large differences in grade participation patterns, curricula and teaching practices across countries, progress is ultimately fairly consistent across the 15 African countries in question. The exceptional across-grade gain for Benin seen in Figure 35 is largely a manifestation of fairly low participation rates among older children – see Figure 33 – and does not reflect an exceptionally successful schooling system in a holistic sense, according to Figure 39. Instead, Gambia emerges as being relatively successful at raising skills between ages 7 and 14. On the other hand, what stands out about Nigeria is that *both* in terms of grade and age, the country appears to do somewhat worse than other countries at advancing the skills of children. Even Tunisia, when compared to São Tomé and Príncipe, emerges as a country which could be expected to do better in terms of the accumulation of children's skills over time.

Figure 39: Gain between age 7 and age 14



Note: The trendline is logarithmic.

To conclude, countries that display relatively high levels of proficiency at the end of primary also do so in the early grades. Differences *across* countries in the proficiency of learners in grade 1, or children aged 7, are very large, and on the whole larger than the gains children make *within* individual countries. This raises important questions around the extent to which factors such as nutrition in the very early years play a role, as opposed to a 'burst' of skills generation in the very first year of so of schooling. Evidence-based answers to these questions in the African context remain scarce.

6 The challenge of measuring improvement over time

The SDGs focus primarily on improvement within countries, and not directly on across-country comparability. Two large African programmes have focussed strongly on both comparability over time and across countries, namely SACMEQ and PASEC. In the case of SACMEQ statistics are designed to be comparable over the following four years: 2000, 2007, 2013 and 2021. The eight countries for which mean scores are easily available for all years are shown in the following two graphs – the greatest hurdle was 2021, a year for which there are few publicly available country reports, though the South Africa national report contains mean results for the eight countries[86].

The fact that for both subjects and for most countries the improvement should have been so much better in the 2007 to 2013 period compared to the earlier 2000 to 2007 period raises questions, especially as comparably dramatic improvements were not seen in PASEC countries for the 2014 to 2019 period – see Figure 42 and Figure 43. All of the eight SACMEQ countries improved by more than 6 SACMEQ points a year, roughly 0.06 standard deviations, in at least one subject. Yet improvements of this speed are seldom observed[87]. In PASEC, three of ten countries exceeded the 0.06 standard deviation annual improvement threshold, which though less striking would still point to remarkable improvements. In the case of SACMEQ, the fact that technical documentation and the microdata beyond 2000 are mostly unavailable means questions around the reliability of the trends are difficult to answer.

Figure 40: SACMEQ trends in mathematics

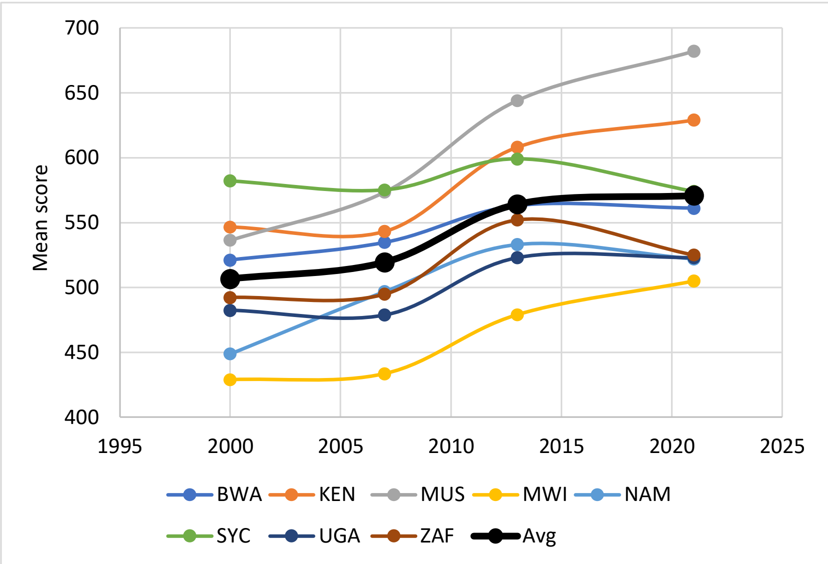
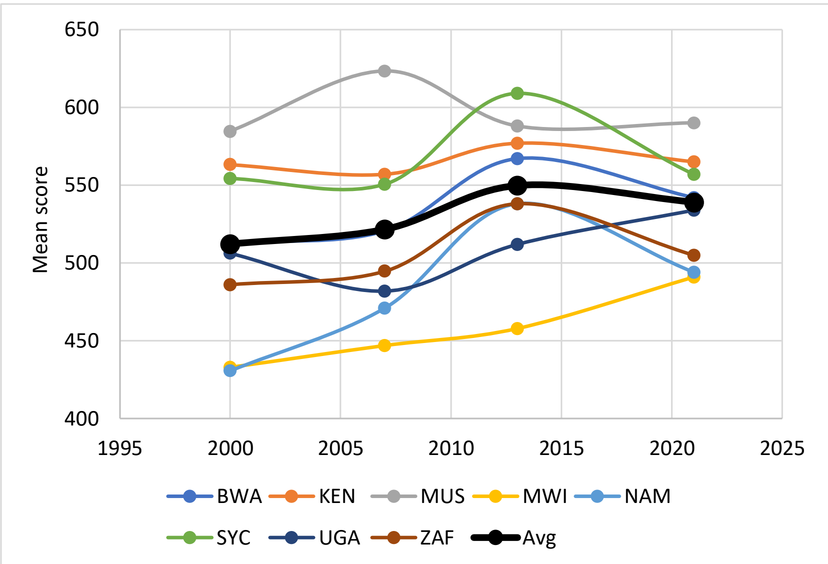


Figure 41: SACMEQ trends in reading



86 South Africa: Department of Basic Education, 2024b. Pre-2013 results are from Makuwa (2010).
87 UNESCO Institute for Statistics, 2019.

The PASEC trends are easier to defend because technical documentation and the microdata are available, yet even here questions have been raised about the national representativity of some samples, based on analysis of access to electricity according to the PASEC learner background questionnaires[88].

Figure 42: PASEC end of primary trends in mathematics

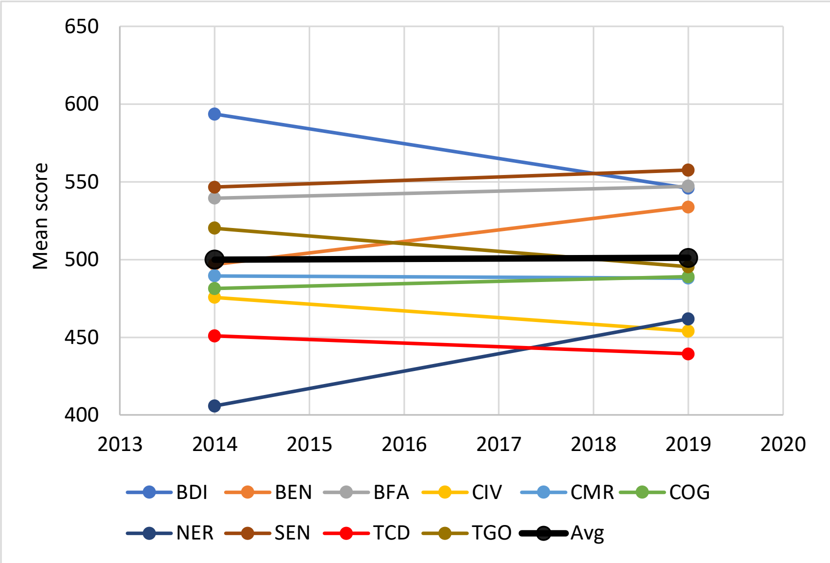
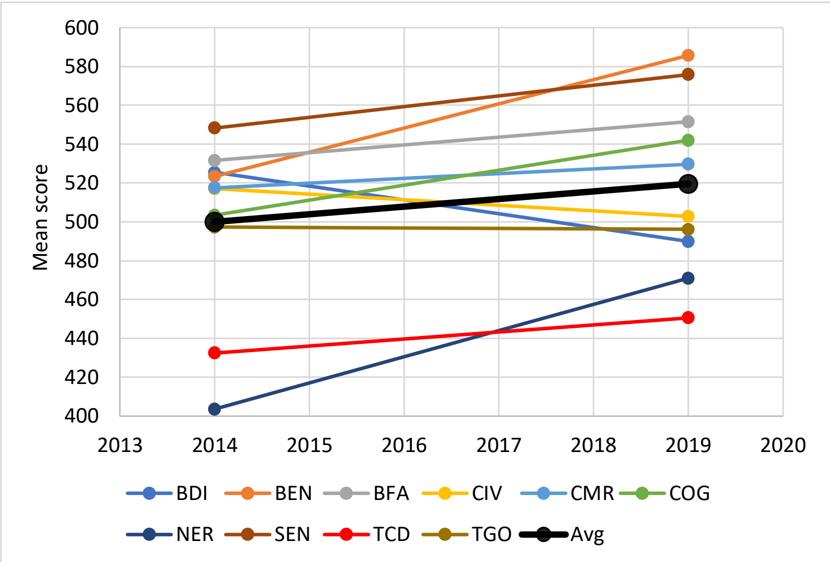


Figure 43: PASEC end of primary trends in reading



It is very noteworthy that despite apparently remarkable improvements seen in SACMEQ and to some extent PASEC, this has received so little attention in government documents in the affected countries. For example, the 2018 to 2022 medium term plan of the Kenyan Ministry of Education[89] makes no mention of SACMEQ (or SEACMEQ) and how this programme has indicated that Kenya has in the recent past been able to achieve remarkably rapid improvements in learning. Nor is there mention of the SDG 4.1.1 indicators dealing with progress in levels of

[88] UNESCO Institute for Statistics, 2021b.
[89] Kenya: Ministry of Education, 2018.

learning proficiency. Even the Kenyan Government SDG progress report of 2019, in its education section, focusses largely on access issues, and pays no attention to trends in learning proficiency. This is not unusual. Government plans and reviews across the continent, and possibly beyond, often fail to focus on progress in terms of comparable measures of learning outcomes.

Arguably, these reporting processes downplay proficiency indicators as these indicators are still relatively new, are technically complex, and often difficult to interpret. Conversations with education planners moreover suggest that the disappointing nature of the proficiency statistics in terms of *levels* of learning easily detract from the importance of improvements. Put differently, even relatively good improvements appear too slow to be interesting for governments. This perception is easily bolstered by the implication of the SDG goals that by 2030 quality schooling for all should be achieved, something which would involve unprecedented speeds of qualitative improvements in schools across the world.

South Africa's 2023 national SDG report[90] is probably unusual insofar as it does make reference to progress in terms of standardised measures of learning proficiency. However, the report avoids reference to the country's SACMEQ results in explaining progress over the post-2000 period, although SACMEQ points to substantive gains. Instead, the report draws only from PIRLS and TIMSS results. This could reflect the higher status of these international programmes, which are especially transparent with respect to their methodologies and microdata.

The 2024 McKinsey report[91] on progress in schooling systems across the world is interesting insofar as of the three developing countries said to be 'sustained & outsized improvers' with respect to learning outcomes in the new century, two are African: South Africa and Morocco (the third country is Peru). Though the report draws from SACMEQ data in identifying Kenya and Malawi as 'emerging improvers', the first choice within the McKinsey methodology is data from the large international programmes TIMSS, PIRLS and PISA.

Raising the status of programmes such as PASEC and SACMEQ would improve the possibility that gains in learning outcomes in African countries are recognised, not just within the continent but also worldwide. It would also improve the sustainability of these programmes. Raising their status would mean ongoing investments to ensure these programmes enjoy skilled analysts and implementers, but also investment in high quality technical products, particularly metadata and microdata. It is worth reminding stakeholders on the continent that not just having educational improvements, but also having indisputable evidence of this, improves the attractiveness of countries for local and international investors, for whom a skilled workforce is important.

7 Conclusion and pointers for future work

The extent to which the current paper takes forward the work of understanding and harmonising Africa's learning statistics can in part be gauged by countries covered, and the recency of the data used. What has this paper added relative to two valuable preceding harmonisation exercises discussed extensively in this paper: Angrist *et al* (2021) and Gust *et al* (2024)?

Angrist *et al* provided proficiency statistics for reading for 94% of the population-weighted continent. Gust *et al* did so for only half of the continent, if imputations of learning levels using non-assessment sources are excluded (section 3.1). The current paper produces across-subject average measures of learning proficiency for countries representing 97% of the continent's children. In Angrist *et al* the most recent Africa data used are from 2017. In what has been presented in the current paper, 26 of African 47 countries draw from data which are from the

[90] Statistics South Africa, 2023.
[91] McKinsey & Company, 2024.

years 2018 to 2023, representing a significant refreshing of the data. This is made possible largely due to the 2019 PASEC results and data having been disseminated, and the availability of MICS foundational learning data, LaNA and AMPL, three sources not available before 2018.

There will almost certainly be future harmonisation exercises, as new data become available. What pointers can the current paper provide for this future work?

Firstly, it seems important to let the analysis be informed by a clear understanding of the users of the final statistics. Education specialists and economists often have different needs, with the former wanting subject-specific learning outcomes measures and the latter often being satisfied with general across-subject measures. Insofar as education specialists are the intended audience, those focussing on specific countries have an interest in the ranking of their country, but especially the possibility of progress over time. As more data become available, the analysis should increasingly evaluate trends over time. Economists are in part interested in measures of learning because of their important relationship with economic progress. Acknowledging this link can help those in education who collect and analyse learning data appreciate the wider importance of their work, and the risks of not having good learning data.

Secondly, harmonisation exercises provide an opportunity to explore *why* certain countries display better proficiency, or better progress over time. Household data such as MICS or PAL are especially valuable in this regard.

Thirdly, an overly narrow view of what data are sufficiently reliable to warrant analysis and what a sufficiently rigorous harmonisation methodology looks like should be avoided. Even data generally considered the best available can display sampling or other problems, as discussed in this paper. In short, there is no perfect data. While the traditional microdata-based approach of using linking countries for harmonisation of programmes is valuable, the paper confirms this does not always render plausible results (see section 3.6.1). This strengthens the case for considering international statistics in the absence of microdata, even if this is not the ideal and microdata ought to be available. The important thing is for any analysis to be frank about the methodological limitations and the reliance on judgement. A valuable by-product of considering data and statistics which are not ideal is that this can yield important information on how the data collection programmes in question could be strengthened.

Fourthly, whether measures of learning should be imputed using non-assessment data is a question that is often asked. One danger of this is that is removes pressure to ensure that all countries do assess learning.

Fifthly, the role of national, as opposed to international, assessments is an important matter. Where international programmes exist they are clearly preferable, as they are specifically designed to facilitate across-country comparability, and can benefit from the pooling of resources. In producing the current paper no national assessment could be found which would add value to the harmonisation. This is despite the fact that some national assessments do attempt to produce internationally comparable statistics (see section 3.6.4). What is required for a national assessment to be comparable to some international programme seems under-appreciated[92].

[92] UNESCO's most recent version of its inventory of learning assessments is in many respects valuable. However, it clearly over-states the ability of assessments to produce statistics which are comparable over countries and time. To illustrate, 544 subject-specific national assessments conducted in the past in Africa are listed in the inventory, of which 260 are said to be of a high quality due to their use of item response theory (IRT). IRT would indeed enhance comparability, yet many assessments of the 260 are well known not to use IRT and are unlikely to produce comparable statistics. See https://tcg.uis.unesco.org/inventory-of-learning-assessments (accessed July 2025).

Turning to what the paper suggests in relation to the utility of existing systems, what stands out is the problem of microdata not being widely available to researchers. A likely reason for the non-availability of microdata is that programmes lack the resources to prepare the data and accompanying manuals for public release. This resourcing problem should be addressed. However, discussions with people involved in the process also suggest that there is sometimes a perception that making the data available increases the risk that errors in the published statistics will be exposed, leading to harm to the assessment programme's reputation. While there is some truth to this, the reputational damage caused by a lack of transparency is likely to be far greater. Conversely, quality control by a wide range of users of the microdata, including researchers not employed by the programme, helps to enhance a programme's status. It is worth remembering that even programmes widely considered to be very reliable have in the past been the subject of important critiques focussing on comparability problems[93].

UNESCO's systems that define proficiency indicators could be enhanced. To illustrate, the current paper has made assumptions around the logic of having TIMSS grade 4 results indicate learning in around grade 6 (section 2.2). There is a plausible logic for this, but it would be beneficial for this to be made more explicit.

The argument is sometimes made that proficiency criteria are set too high for Africa, resulting in discouragingly low values – the median schooling system in Africa sees only around 11% of learners being proficient (Figure 30). This argument is compelling – even Finland sees only around 75% of end of primary learners reaching the minimum threshold in mathematics using existing criteria[94]. Yet changing standards that have been in existence for years, and are the result of extensive work, is not easy. Using less stringent thresholds, such as the ability of a child to read a sentence out loud, is valuable, but so is measurement using well-established global thresholds. In either case, rankings across countries are likely to be similar. Moreover, if progress is occurring, either monitoring route should pick this up.

[93] Jerrim, 2013.
[94] See footnote 78.

References

Angrist, N., Djankov, S., Goldberg, P.K. & Patrinos, H.A. (2021). Measuring human capital using global learning data. *Nature*, 592: 403-408.

Awich, M. (2021). *The SACMEQ IV project in international: A study of the conditions of schooling and the quality of education.* Gaborone: SACMEQ.

Brombacher, A., Bulat, J., King, S. *et al* (2015). *National assessment survey of learning achievement at grade 2: Results for early grade reading in Zambia*. Research Triangle Park: RTI International.

Cambridge Education (2019). *Evaluation of the back to learning initiative in South Sudan: Final report*. Cambridge.

CONFEMEN (2014). *Cadre de référence des test PASEC 2014 de langue et de mathématiques de debut de scolarité primaire.* Dakar.

CONFEMEN (2017). *Manuel d'exploitation des données évaluation internationale PASEC 2014*. Dakar.

CONFEMEN (2020). *PASEC 2019: Qualité des systèmes éducatifs en Afrique subsaharienne francophone*. Dakar.

CONFEMEN (2022). *Rapport technique de l'évaluation internationale PASEC2019*. Dakar.

Cortez Ochoa, A.A. & Sandoval Hernandez, A. (2023). *Feasibility of using the data produced by the Early Grade Reading (EGRA) and Early Grade Mathematics (EGMA) to measure and monitor SDG 4.1.1 by complementing it with other banks of ítems*. Montreal: UNESCO Institute for Statistics.

Dharamshi, A., Barakat, B., Alkema, L. & Antoninis, M. (2021). *Adjusted Bayesian completion rates (ABC) estimation*. Toronto: University of Toronto.

Education Development Center (2017). *USAID Literacy, Language and Learning initiative (L3): National fluency and mathematics assessment of Rwandan schools: Endline report*. Waltham.

Gove, A. & Wetterberg, A. (2011). The Early Grade Reading Assessment: An Introduction. In Amber Gove and Anna Wetterberg (ed.), *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park: RTI Press.

Gust, S., Hanushek, E.A. & Woessmann, L. (2024). Global universal basic skills: Current deficits and implications for world development. *Journal of Development Economics*, 166.

Gustafsson, M. (2013). More countries, similar results: A nonlinear programming approach to normalising test scores needed for growth regressions. *Studies in Economics and Econometrics,* 37(2): 93-111.

Gustafsson, M. & Barakat, B. (2023). Marginalized learners in international and regional test data: The extent of floor effects. *Comparative Education Review*, 66(3): 509-533.

Hill, C.J., Bloom, H.S., Black, A.R. & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives,* 2(3): 172-175.

Jerrim, J. (2013). The reliability of trends over time in international education test scores: Is the performance of England's secondary school pupils really in relative decline? *Journal of Social Policy,* 42(2): 259-279.

Kenya: Ministry of Education (2018). *National education sector strategic plan for the period 2018-2022*. Nairobi.

King, S., Korda, M., Nordstrum, L. & Edwards, S. (2015). *Liberia Teacher Training Program: Endline assessment of the impact of early grade reading and mathematics interventions*. Research Triangle Park: RTI International.

Makuwa, D.K. (2010). Mixed results in achievement. *IIEP Newsletter,* XXVIII(3).

Mauzunya, M. & Varly, P. (2012). *Evaluation des compétences fondamentales en lecture au Burundi: Rapport final provisoire*. Paris: SOFRECO.

McKinsey & Company (2024). *Spark & sustain: How all the world's school systems can improve learning at scale*. New York.

Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Chestnut Hill: Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L. & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Chestnut Hill: Boston College.
Mullis, I.V.S., Von Davier, M., Foy, P., Fishbein, B., Reynolds, K.A. & Wry, E. (2023). *PIRLS 2021 international results in reading*. Chestnut Hill: Boston College.
PAL Network (2020). *International Common Assessment of Numeracy: Background, features and large-scale implementation.* Nairobi.
PAL Network (2021). *Out of school children: Leave no one behind in education.* Leeds: University of Leeds.
Patrinos, H.A. & Angrist, N. (2018). *Global dataset on education quality: A Review and Update (2000–2017).* Washington: World Bank.
Piper, B. (2010). *Ethiopia Early Grade Reading Assessment: Data Analysis Report: Language and Early Learning.* Research Triangle Park: RTI International.
Pouezevara, S., Sock, M. & Ndiaye, A. (2010). *Evaluation des compétences fondamentales en lecture au Sénégal: Rapport d'analyse*. Research Triangle Park: RTI International.
RTI International (2014). *Egypt grade 3 early grade reading 2nd national assessment*. Research Triangle Park.
RTI International (2015). *Early Grade Reading Assessment (EGRA) national baseline assessment in Mali: Report of findings*. Research Triangle Park.
Rwanda: National Examination and School Inspection Authority (2023). *Learning achievement in Rwandan schools analysed using global proficiency benchmarks.* Kigali.
Sandefur, J. (2016). *Internationally comparable mathematics scores for fourteen African countries.* Washington: Center for Global Development.
Singh, A. (2019). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, 18(4): 1770–1813.
South Africa: Department of Basic Education (2024a). *South African Systemic Evaluation 2022: Volume 1: Technical report*. Pretoria.
South Africa: Department of Basic Education (2024b). *SEACMEQ V South African technical report: A study of the conditions of schooling and the quality of education*. Pretoria.
Statistics South Africa (2023). *Sustainable Development Goals country report 2023 South Africa*. Pretoria.
UNESCO Institute for Statistics (2017). *Counting the number of children not learning: Methodology for a global composite indicator for education*. Montreal.
UNESCO Institute for Statistics (2019). *How fast can levels of proficiency improve? Examining historical trends to inform SDG 4.1.1 scenarios*. Montreal.
UNESCO Institute for Statistics (2021a). *Minimum Proficiency Levels used to report for indicator 4.1.1.* Montreal.
UNESCO Institute for Statistics (2021b). *Assessing learning proficiency levels and trends for Sustainable Development Goal 4.1: A focus on Africa.* Montreal.
UNESCO Institute for Statistics (2021c). *Continental Overview: Bridging CESA and SDG 4 in Africa.* Montreal.
UNESCO Institute for Statistics (2022a). *COVID-19 in Sub-Saharan Africa: Monitoring impacts on learning outcomes.* Montreal.
UNESCO Institute for Statistics (2022b). *Analysis report: Establishing a concordance between PASEC and TIMSS/PIRLS.* Montreal.
UNESCO Institute for Statistics (2024a). *Trends in learning proficiency in the last twenty years: How close are we to reliable regional and global SDG 4.1.1 trend statistics?* Montreal.
UNESCO Institute for Statistics (2024b). *Assessments for minimum proficiency levels a and b (AMPL-ab): International report.* Montreal.
UNESCO Institute for Statistics (2024c). *GAML/TCG criteria for use of an assessment to report on SDG 4.1.1: Draft 2, 25 March 2024.* Paris.
UNICEF (2020). *Toward achieving inclusive and equitable quality education for all: A manual for statistical data analysis using Multiple Indicator Cluster Surveys (MICS6) with a special focus on achieving the Sustainable Development Goals.* New York.

UNICEF (2024). *The state of the world's children 2024: The future of childhood in a changing world: Statistical compendium.* New York.
USAID (2016). *Early Grade Reading Assessment (EGRA) toolkit*. Washington.
Uwezo (2017). *Are our children learning? Uwezo Tanzania sixth learning assessment report*. Dar es Salaam.
Von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., *et al* (2024). *TIMSS 2023 International Results in Mathematics and Science*. Chest: Boston College.
Von Davier, M., Khorramdel, L., Reynolds, K., Aldrich, C.E.A. *et al* (2025). *LaNA 2023 linking study results*. Chestnut Hill: Boston College.
Woldehanna, T., Behrman, J.R. & Araya, M.W. (2017). The effect of early childhood stunting on children's cognitive achievements: Evidence from young lives Ethiopia. *Ethiopian Journal of Health Development*, 31(2): 75-84.
World Bank (2019). *Implementation completion and results report on a grant in the amount of SDR 55.4 million (US$76.5 million equivalent) to the Sudan for the Sudan Basic Education Recovery Project*. Washington.

AFLEARN
DATAHUB

African Foundational Learning Data Hub
DataFirst
University of Cape Town
www.datafirst.uct.ac.za/aflearn