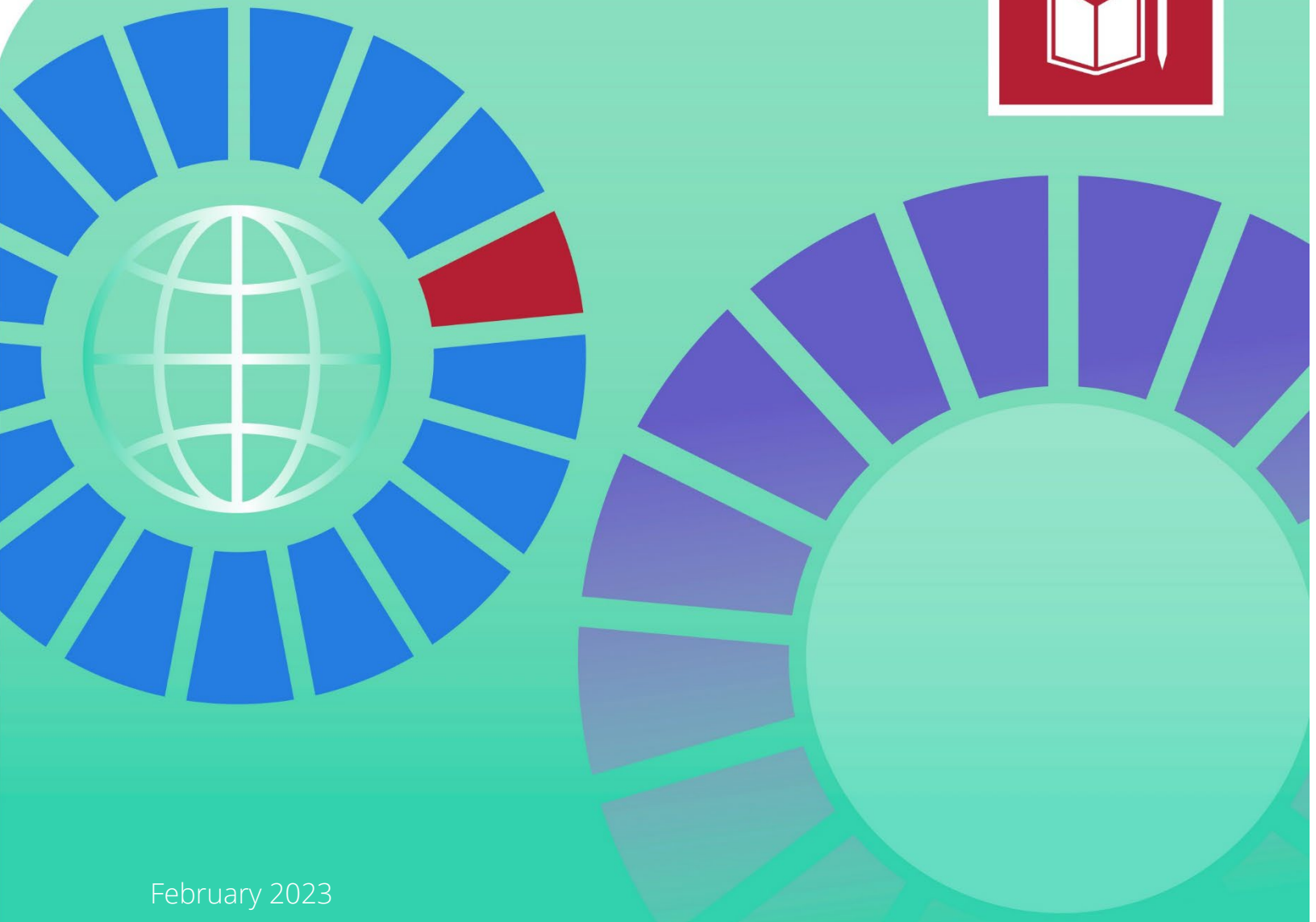
unesco
Institute for Statistics



February 2023

Trends in learning proficiency in the last twenty years:

How close are we to reliable regional and global SDG 4.1.1 trend statistics?

UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfil its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical cooperation to Member States for their development policies and projects; and 5) the exchange of specialized information.

UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science, technology and innovation, culture and communication. The UIS was established in 1999. It was created to improve UNESCO's statistical programme and to develop and deliver the timely, accurate and policy-relevant statistics needed in today's increasingly complex and rapidly changing social, political and economic environments.

Published in 2023 by: UNESCO Institute for Statistics

Report prepared for the UNESCO Institute for Statistics by Martin Gustafsson
mgustafsson@sun.ac.za
Research on Socio-Economic Policy (ReSEP)
University of Stellenbosch

C.P 250 Succursale H
Montréal, Québec H3G 2K8
Canada
Email: uis.publications@unesco.org
http://www.uis.unesco.org

978-92-9189-320-1

UIS/2023/ED/IP/65

© UNESCO-UIS 2023

This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (http://creativecommons.org/licenses/by-sa/3.0/igo/). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (http://www.unesco.org/open-access/terms-use-ccbysa-en). The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities or concerning the delimitation of its frontiers or boundaries. The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

SUMMARY

The milestone 2005 UNESCO Global Monitoring Report, titled *Education for all: The quality imperative*, marked an important shift towards measuring the learning proficiency of children, not just their attendance in school. The emphasis on proficiency is clearly present in the Sustainable Development Goals, introduced in 2015. SDG Indicator 4.1.1 requires the monitoring of reading and mathematics proficiency at three school levels. The 2030 target put forward by countries was 100% proficiency for each of these six indicators.

The current report focuses on estimating regional and global trends for the six 4.1.1 indicators during the pre-pandemic period 2000 to 2019. The impacts of the pandemic on proficiency have been devastating and exceptional and are covered in several existing reports. Understanding past proficiency trends, including those from before the pandemic, is of obvious importance for understanding what is possible in the years up to 2030. This understanding can assist in decisions around how adequate or inadequate past efforts to improve learning have been.

The 100% targets for 2030 require global gains to be around 2.7 percentage points a year between 2015 and 2030 – this assumes a starting point of around 60% proficiency in 2015, which is based on actual patterns[1]. **Past estimates of global annual gains[2], based on less comprehensive methods than those used for the current report, have ranged from positive 0.7 to *negative* 1.0. Clearly, this would not put the world on track to reach the SDG 4.1.1 targets.** The current report confirms that reaching the 2030 targets would be extremely difficult. At the same time, gains in the 2000 to 2019 period appear to have been *positive*, though small.

Data sources from the UIS

The analysis presented here draws from 1,684 proficiency statistics available in the UIS.Stat reporting system for the 2000 to 2019 period, which produce 1,097 two-point trends. It also draws from UIS.Stat statistics on participation in schooling. The latter statistics are important insofar as the practice has been for Indicator 4.1.1 statistics to cover just children *within* schooling systems, not those who are out of school. This practice has come about more due to the way data are collected than to any deliberate decision around definitions. **Moreover, the relatively high availability of *participation* statistics underscores the fact that *proficiency* statistics should become even more available than is currently the case.** The current analysis also uses World Population Prospects data. The share of the world's children accounted for by individual countries and regions is continually changing, and this must be taken into account when national indicator values are aggregated.

More data needed to determine proficiency trends

Though this report arguably represents the most comprehensive attempt to date to estimate regional and global long-range proficiency trends, this does not mean that the aggregate trends presented in the report are highly reliable. **While only 6% of the world's children are in countries where there**

1 Figure 19.
2 Section 9.

are no proficiency statistics at all, at any level, 52% of children are in countries where statistics exist, but there is no series of comparable values, meaning no trends are detectable[3].

In some regions, trends for the critical early grade reading indicator are mostly absent. For instance, in sub-Saharan Africa only 29% of children are in countries which *do* measure this[4]. Other regions, such as Central and Southern Asia, have no trend data for this indicator at all.

Much of the lack of trend data is accounted for by India and China, in which 31% of the world's children live and which have no trends for 2000 to 2019 in the published SDG statistics, for any level or subject. However, there are an additional 104 countries which also display the same absence of trend data.

A key reason why the just over half of countries which do have trend data cannot be considered representative of all the world's countries is that countries without data tend to be countries with more serious governance problems. This suggests that these excluded countries would be *less* likely to display proficiency gains. Put differently, the presence of trend data is not randomly spread across the countries of the world. This pattern is true even if world regions are analysed separately[5].

Available data show modest gains in learning proficiency

Though the available statistics are likely to over-state the gains made, these statistics point to very modest gains for the 2000 to 2019 period, relative to what is needed to reach the 2030 targets of 100%. **In reality, annual improvements in the 2000 to 2019 period have optimistically been around 0.39 percentage points a year, or just one-seventh of the necessary 2.7 annual gain needed for 100% proficiency**. This figure of 0.39, which refers to end of primary reading[6], is optimistic in the sense that it assumes that where trends have been detected in countries, the years during which measurement took place were randomly selected, meaning countries were not more likely to measure change during periods when improvements were more likely to be occurring. The assumption is optimistic because in reality one could expect better measurement to be occurring when there is a strong focus on improvement initiatives.

Because participation in schooling has improved substantially over the 2000 to 2019 period, it is important to take this into account when examining proficiency trends. It is typically assumed that children who do not complete specific levels of schooling do not achieve the proficiency expected at those levels. Using this assumption and a new harmonized dataset of school completion, it is estimated that the abovementioned 0.39 percentage point gain for end of primary reading becomes 0.53 after completion of primary schooling is considered[7]. This is in a context where primary school completion has improved from 77% in 2000 to 86% in 2019[8]. **Improvements in primary school completion have thus helped to improve levels of proficiency among children.** Even so, annual gains have remained around one-eighth of what they would need to be to reach the 100% targets by 2030.

[3] Table 8.
[4] Table 9.
[5] Section 6.
[6] Table 9.
[7] Table 13.
[8] Figure 20.

An annual gain of 0.53 percentage points a year translates to just 5.3 percentage points over ten years. While this may seem unsatisfactory, the historical reality represented by the 0.53 annual gain statistic is important, for two key reasons. Firstly, it suggests that a country achieving this rate of improvement is not failing, it is simply achieving what has *on average* been achieved in the past. Any improvement above this level would be *above average*, and arguably satisfactory, at least relative to past trends. This would be a conservative assumption as the 0.53 annual gain statistic is likely to be an over-estimate, both because it ignores countries with worse governance systems and no data, and the fact that countries which do measure trends are more likely to do so when improvements are occurring. Secondly, the 0.53 average annual gain underscores how rigorous systems measuring proficiency need to be. Even small changes can be significant.

The end of primary reading indicator receives attention here both because it is an important reflection of the learning process, and because the available data behind this indicator cover a greater proportion of the world's children, compared to the other five SDG 4.1.1 indicators. Specifically, 34% of the world's children are represented by the end of primary reading statistics[9]. This drops to 32% if both proficiency and participation statistics are used[10].

Turning to the other five indicators, all except for end of primary mathematics produce positive annual gains. The mean across the five positive annual gain statistics is 0.4 (before improvements in school completion have been taken into account), and none of the five reach 1.0 percentage point a year[11]. This is at the global level.

The abovementioned 0.53 percentage point annual gain for end of primary reading is largely driven by very high annual gain statistics of over 2.0 in the two regions Central and Southern Asia and Eastern and South-eastern Asia. The statistics for these two regions must clearly be interpreted with much caution as they draw from just 12% and 16% of the regional child population respectively. Just one other region in the developing world, sub-Saharan Africa, displays a positive annual gain for end of primary reading. Here the gain of 0.14 percentage points a year is relatively representative, as it is based on 39% of the child population[12].

When trends are examined separately for males and females[13], the rates of progress often differ slightly by sex, but the differences are small and no obvious patterns emerge. What is noteworthy with respect to the *level* of proficiency, is that females tend to outperform males in reading, while males tend to outperform females in mathematics.

Steps to improve monitoring of proficiency trends

What can be done to improve the monitoring of regional and global proficiency trends in the coming years? **Clearly, much depends on the capacity and willingness of countries to become more deeply involved in this type of monitoring, be it in the form of international testing programmes or national programmes.** There is much that global bodies, such as the UNESCO Institute for Statistics (UIS), can do to promote this. Engagement with individual countries or individual cross-country testing programmes, such as PASEC, SACMEQ and LLECE, with a view to building

[9] Table 9.
[10] Table 13.
[11] Table 9.
[12] Table 13.
[13] Section 12.

capacity and promoting the availability of high-quality microdata and metadata, is important. Rigorous analyses of the data by the UIS can be considered capacity building insofar as this provides, for instance, national analysts with methods and insights. Clearer procedures whereby countries report on proficiency to the UIS, through a bottom-up approach, can result in more national ownership of statistics, and also more reliable trends. This can complement the existing top-down approach, whereby around 97% of proficiency statistics in UIS.Stat are drawn from international programmes. International programmes are very valuable, in part because they can enhance measurement rigour and comparability across time and countries, and in part because they are less subject to undue interference by national authorities, in a context where proficiency trends are inevitably sensitive from a political perspective. At the same time, promoting national programmes, which can be tailored to local needs, is important too.

Clearly one thing that is urgently needed is a more standardized approach to viewing Indicator 4.1.1 statistics in relation to participation statistics. The two should not be viewed in isolation. Yet how best to bring them together is not a straightforward matter. This report has provided proposed methods for addressing this.

Contents

1. Introduction

National and international debates around progress in schooling systems have over many years shifted towards an emphasis on how well children learn, not just attendance in school. This shift began two decades ago, with the 2005 UNESCO Global Monitoring Report[14], titled *Education for all: The quality imperative*, marking an important milestone. Learning proficiency subsequently became a central feature of Goal 4 within the Sustainable Development Goals (SDGs), introduced in 2015.

There is a clear demand for regional and global statistics relating to the SDGs, including Indicator 4.1.1, which deals with the 'proportion of children and young people … achieving at least a minimum proficiency level in (i) reading and (ii) mathematics'. Mandatory disaggregation of this indicator is by level: grades 2/3; end of primary; and lower secondary. This produces six indicators. In addition, each of the six should be disaggregated by sex. The 2030 target countries put forward called for 100% of children achieving proficiency in each of the six indicators.

Reporting bodies in the UN system have responded to this demand with a variety of estimates in published reports. Producing regional and global statistics is clearly among the responsibilities of the UIS[15]. However, reporting on Indicator 4.1.1 *trends* at the regional and global levels remains infrequent and methodologically problematic, as will be explained in this report.

During the COVID-19 pandemic, global and regional estimates of learning losses arising from the disruptions to schooling have been widely quoted. Global and regional trends in this regard have become a matter of interest to a degree not seen before the pandemic[16]. What is needed is to sustain a focus on global, but also regional, proficiency trends, through rigorous methodologies and practices. Global estimates produced during the pandemic were a response to a global emergency, and had to be put together hastily. An opportunity now exists to learn from past attempts and to plan for more reliable estimations of trends.

Two key research questions inform this report:

1. What do we know and not know about regional and global SDG 4.1.1 trends since around 2005, up to the start of the pandemic in 2020?

2. What are the critical factors needed to improve the availability of regional and global SDG 4.1.1 statistics in the years up to 2030?

Section 2 below introduces the statistics currently available on the UIS.Stat platform needed for learner proficiency trends. **Section 3** discusses the reliability of the trends currently seen, with reference to unusual outlier trends and the consistency of data sources. **Section 4** explains confidence intervals arising from the sample-based nature of most learning proficiency datasets, and what this means for understanding aggregate regional and global trends.

Section 5 discusses the extent to which international testing programmes, which account for 97% of the proficiency statistics available in UIS.Stat, make their microdata and metadata available to a range

[14] UNESCO, 2005.
[15] UNESCO, 2017.
[16] UIS, 2022a; Azevedo *et al*, 2022; Moscoviz and Evans, 2022.

of analysts. The more such materials become available, and the more analysts use them, including analysts based in national institutions, the greater the certainty around data quality.

Section 6 examines statistically which types of countries have and do not have SDG 4.1.1 statistics pointing to proficiency levels and trends. Currently, just 5% of school-age children are in countries where there are no proficiency statistics, at any schooling level. However, 52% of school-age children are in countries for which a *trend* cannot be established for the two decades preceding the pandemic through the available UIS statistics. Moreover, the monitoring of trends in developing countries with respect to the acquisition of reading in the early grades, an especially critical indicator and a prerequisite for subsequent schooling, is often weak. For instance, in sub-Saharan Africa there is essentially no such monitoring for 71% of children over two decades.

Section 7 looks at how the concentration of the population in specific countries per world region affects the ease with which regional averages can be estimated. **Section 8** introduces national and regional child population trends, a matter that can influence global trends to a large degree.

Section 9 presents the methods used in the current report to aggregate national proficiency statistics into regional and global trends. A clear limitation is the low availability of trend statistics, especially in certain regions and for specific schooling levels. Moreover, countries without trend statistics appear to be countries with larger political and social problems, or with less organized schooling systems, suggesting that existing statistics will produce overly positive aggregate trends.

Sections 10 and **11** present aggregate trends where both sexes are combined. Even with the risk that aggregate trends will be overly positive, regional and global proficiency trends are remarkably flat for the period 2000 to 2019. However, this should be seen against the fact that *participation* in schooling, at least at the primary level, has steadily improved. This suggests strongly that although official SDG 4.1.1 statistics, which refer only to children *attending* school have not improved markedly, proficiency levels in the *child population* have improved.

Section 12 examines what the statistics say about differences in trends across gender.

Section 13 presents a basic methodology for combining proficiency and participation statistics with a view to obtaining estimated trends for proficiency in the *child population.*

Finally, **Section 14** addresses the question of how the availability of regional and global trends for Indicator 4.1.1 can be improved. Clearly, this is largely about addressing gaps at the national level, in particular those pertaining to China and India, which jointly account for 31% of the world's children. However, there are further approximately 104 smaller countries for which trend statistics also need to be improved. This section also addresses how trends are reported, and specifically the need for a more systematic integration of proficiency statistics emerging from testing programmes and statistics on school participation.

Appendices 1 and **2** provide aggregations, along the lines of the regional aggregations, according to World Bank income category and according to whether a country is in the Global Partnership for Education (GPE) or not. **Appendix 3** provides an analysis of PASEC[17] microdata, in part to demonstrate the utility of such analysis for probing data reliability questions.

[17] Programme d'analyse des systèmes educatifs (Programme for the Analysis of Education Systems).

2. The availability of national trend statistics

This report draws largely from what is labelled the '4 March 2022 Release' Indicator 4.1.1 statistics in the SDG section of UIS.Stat. These data were downloaded in September 2022. The 228 countries reflected in **Table 1** are those found in the UIS.Stat demographic data tables after three tiny countries with missing population values had been removed[18]. Using only statistics that combine both sexes, there are 1,684 proficiency statistics in the 4.1.1 data, spread across 160 countries. Here and elsewhere in the report, any statistics after 2019 are ignored, as the focus is on the pre-pandemic period. Had 2020 and 2021 statistics been included, there would have been 1,696 proficiency statistics. The 1,684 statistics reflected in Table 1 produce 1,097 level- and subject-specific two-point trends spread across 120 countries. To illustrate the meaning of a two-point trend, if a country had proficiency statistics for the end of primary reading for 2009, 2013 and 2017, this would produce two two-point trends: one for 2009 to 2013, and another for 2013 to 2017. The average length of the 1,097 two-point trend is 4.1 years.

These 120 countries with trends are 53% of all countries, and account for 48% of the world's children (this will be seen in **Table 8**).

More than half of the 52% of children in countries with no trend are from China and India, two countries with no trend in the 2000 to 2019 period. These two countries account for 31% of the world's children, the larger of the two being India, with 18% of the world's children. This still leaves 104 countries other than China and India which do not have trend data, at least not for the two decades preceding the pandemic[19].

Table 1: Availability of national trends

	Countries	Number of 4.1.1 level statistics	Countries covered in previous column	Number of 4.1.1 trends	Countries covered in previous column	% with trend
Sub-Saharan Africa	52	215	40	80	29	56
Northern Africa and Western Asia	24	199	21	135	20	83
Eastern and South-eastern Asia	18	143	16	86	8	44
Central and Southern Asia	14	66	11	24	4	29
Oceania	21	59	6	40	2	10
Latin America and the Caribbean	48	305	24	192	17	35
Europe and Northern America	51	697	42	540	40	78
World	228	1,684	160	1,097	120	53

As will be seen in Section 10, around 100% of the world's children are in countries with trend data for *participation* in schooling. This is a key reason why reliable regional and global trends for learning proficiency, or the quality of schooling, are far more difficult to estimate than trends for participation. Put differently, while the SDGs have shifted the emphasis to where it is needed, namely how well

[18] Holy See, Nieu, Tokelau.

[19] While India's ASER programme publishes countrywide statistics on proficiency trends, based on household surveys, these exclude urban centres. ASER, which is considered to measure trends with sufficient reliability – see Johnson and Parrado (2021) – points to basic proficiency improving at the primary level by around half a percentage point a year between 2012 and 2018 (ASER Centre, 2019). But levels are low, the percentage proficient in the second grade improving from 7% to 9% over this period.

children learn, these goals pose a much larger monitoring challenge than the preceding Millennium Development Goals, whose emphasis with respect to schooling was largely on participation.

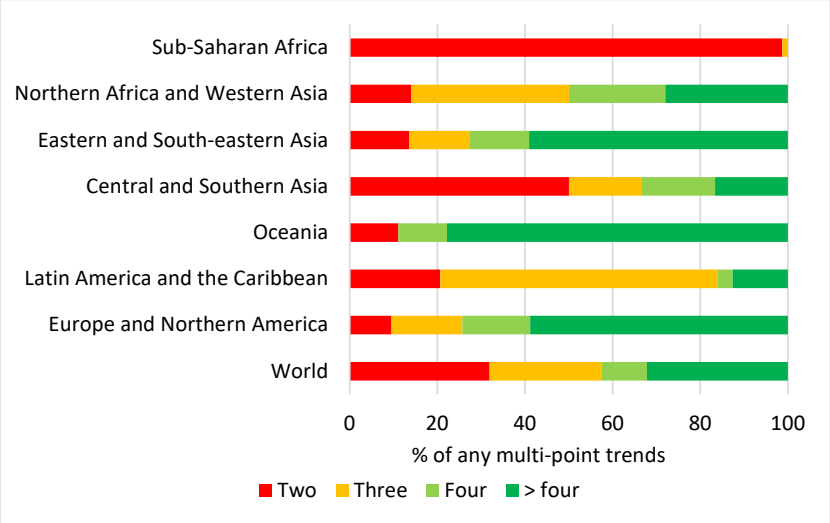
The 1,097 two-point trends referred to previously produce 395 country-level and subject-specific multi-point trends, where for instance points in 2009, 2013 and 2017 would produce *one* (mostly) non-linear trend. Within the 395 would be two-point trends which are not extended beyond the two points. How these 395 trends are divided across regions is shown in **Table 2**. The average number of years covered per trend is also shown. Sub-Saharan Africa clearly lacks longer trends. The ideal would be to have a high percentage of countries with trends (last column of Table 1), but also a high average number of years covered (last column of Table 2). In this regard, apart from the largely developed countries in Europe and Northern America, the region Northern Africa and Western Asia stands out as being relatively well endowed with statistics required to measure trends.

Table 2: Length of all multi-point trends

	Number of level- and subject specific multi-point trends	Average coverage per trend in years
Sub-Saharan Africa	79	5
Northern Africa and Western Asia	50	11
Eastern and South-eastern Asia	22	14
Central and Southern Asia	12	8
Oceania	9	16
Latin America and the Caribbean	87	12
Europe and Northern America	136	14
World	395	11

Figure 1 provides further details regarding the length of multi-point trends, by focusing on the number of points per trend. In sub-Saharan Africa, nearly all trends are just two-point trends.

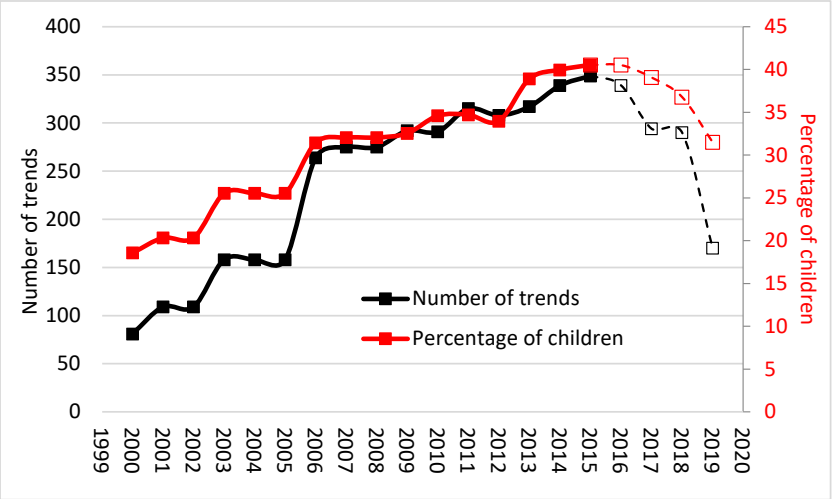
Figure 1: Number of points per trend by region



There is a limited problem of a lack of more recent trends. To illustrate, of the 395 trends seen in Table 2, 40 end before 2015 and five end before 2010. This limitation does not seem too serious for the

purposes of the analysis that follows. As can be seen from the following graph, which adds up the number of trends present in any year, among the 395, there has been an increase in the availability of trend data over time. This is also true if trends are weighted by the child population in a recent year, and each country is counted just once in a year – see the red curve in the graph. (Given that single two-point trends are on average four years in length, the decline in the last four years, 2015 to 2019, can be ignored. During these years it can be assumed that many end-point results have not become available.)

Figure 2: Increasing availability of trend statistics



Note: As an example, a trend for 2010 to 2014 would result in a count of one for each of the years 2010, 2011, 2012, 2013 and 2014 with respect to 'Number of trends'.

3. National trends with some reliability controls

For the purposes of the current report, there are three key reasons why proficiency trend statistics can be unreliable[20]. Firstly, there may be sampling problems. As will be seen, most 4.1.1 statistics are from sample-based testing systems. Apart from the obvious risk that samples may not be nationally representative at different points in time, sampling problems include inconsistencies with respect to when tests are administered. Learners who have just entered Grade 6 and learners who are halfway through Grade 6 are not comparable, even in a context where samples are nationally representative. Changing levels of grade repetition can also influence the comparability of groups of learners from different years. Secondly, what occurs in the testing venue can change over time. A very different test design, different time limits, and a different type or degree of assistance to children by test administrators could apply. Thirdly, there can be problems in the data processing or in the statistical approaches intended to improve the comparability of results.

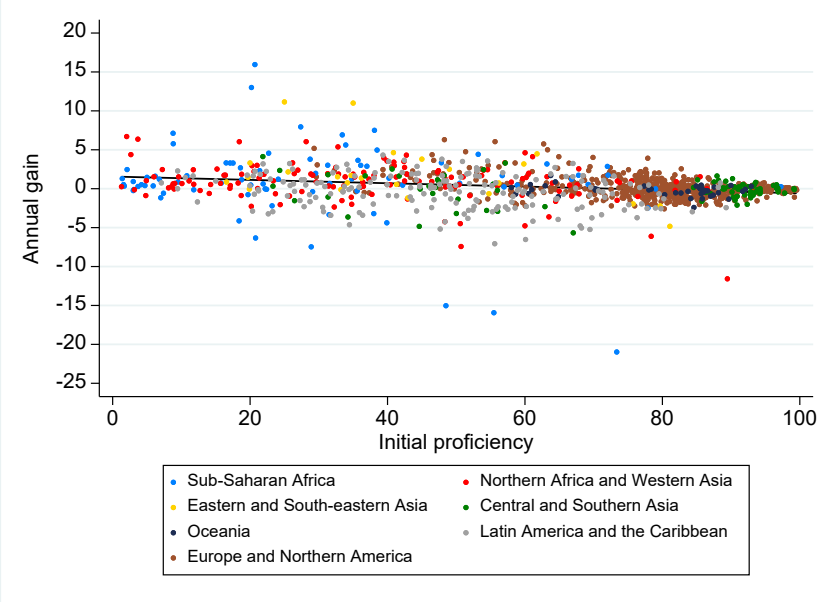
Exceptionally steep upward or downward trends could point to comparability problems flowing from any of the three factors discussed above. **Figure 3** reflects the 1,097 two-point trends referred to in Table 1. Of these trends, 595 trends are greater than zero. None are exactly zero, meaning the remaining 502 are negative. Only slightly more than half of the trends display an improvement.

[20] UIS, 2017a, 2018.

Of the eight annual gains which exceed 10 percentage points in absolute terms, five are from sub-Saharan Africa. It is clear from the graph that both positive and negative outliers from this region are fairly common. As might be expected, trends with a lower initial point tend to be associated with a higher gain, though the differences are small. The trendline points to a typical gain of 1.3 percentage points a year where the point of departure is 10% of learners being proficient, and a loss of 0.4 percentage points where the point of departure is 90% of learners being proficient.

An earlier UIS analysis of trends in Africa concludes that the continent has displayed gains which are exceptionally large, even if the lower point of departure in Africa is taken into account. To some extent, analysis of both the test and background data associated with typical testing programmes can evaluate whether inconsistencies with respect to sampling, test administration and data processing unduly influenced trends. **Appendix 3** examines PASEC microdata and concludes that sampling and the conversion of raw data to trend statistics do not appear to have unduly affected trends in PASEC, though questions with regard to individual countries do arise. It is clear that providing a broad range of analysts access to a testing programme's data can enhance certainty around strengths and current limitations of particular programmes.

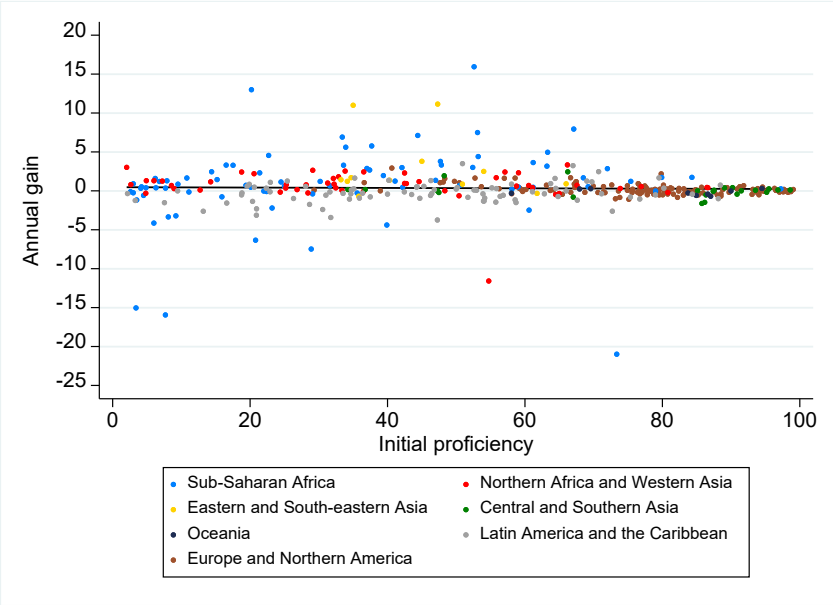
Figure 3: Outlier annual gains among two-point trends



If the 395 multi-point trends reflected in Table 2 are graphed, along the lines of Figure 3, the result is the following Figure 4. Of the 395 trends, 234 or over half are positive and 16 display absolute annual gains (or losses) exceeding five percentage points.

The simple unweighted mean for the 395 points shown in **Figure 4** is a gain of 0.4 percentage points a year. *If the target of proficiency of 100% were to be obtained by 2030, the average annual gain would need to be around 2.7 percentage points.* This assumes a global average level of proficiency of 60% in 2015, the approximate 2015 reality for end of primary reading (**Figure 19**). This represents a speed of improvement around seven times faster than the average seen in Figure 4. Put differently, only 30 of the 395 points in Figure 4 represent annual gains of 2.7 percentage points or more. As already pointed out in UIS (2020), while reaching the 100% targets in 2030 is a noble aspirational goal, this would only be possible if the entire world achieved what has historically been exceptional rates of improvement

Figure 4: Outlier annual gains among 395 multi-point trends



Earlier UIS work suggests that annual gains in percentage proficient statistics are unlikely to exceed around three percentage points[21]. If the assumption were used that a change exceeding five percentage points a year should be considered possibly unreliable, then the number of two-point trends would be reduced from 1,097 to 1,064.

Turning to the consistency of data sources behind the 4.1.1 statistics, UIS.Stat now provides data sources as an easily accessible additional variable when data are downloaded. All of the 1,684 proficiency statistics referred to in Table 1 had some source described in the metadata variable. In all, there were 84 unique metadata descriptions. A normalized set of metadata descriptions were generated, following two analytical steps. First, where original metadata descriptions differed from each other only with respect to the grade (or level) and subject, one common descriptor was generated. Thus, several original descriptors all describing the international PASEC programme were changed to 'PASEC'. Moreover, any national monitoring programme tracking a specific level and subject was given the name of the country. Where a monitoring programme provided statistics for just one year, in other words without producing a trend, the new metadata descriptor was left blank. The exercise of normalizing metadata descriptors, and excluding metadata not associated with a two-point trend, resulted in the ten descriptors seen in **Table 3**. Of the 1,097 two-point trends referred to in Table 1, 1,086 had the same descriptors for both points in time. The 11 cases with inconsistent data were from five African countries: Ethiopia, Kenya, Lesotho, Uganda, and Zambia. For instance, in Kenya SACMEQ was used for end of primary mathematics in 2013, while a national assessment was used for the corresponding monitoring in 2019. In the case of Ethiopia, national assessments were used for both 2016 and 2019 statistics, but these were clearly different in their approach, according to the metadata, and produced very different proficiency statistics. Of the five countries, Kenya and Uganda also had two-point trends where the metadata *were* consistent.

[21] UIS, 2019.

If both reliability controls referred to above, namely the exclusion of overly steep trends and trends with inconsistent sources, the 1,097 two-point trends are reduced to 1,059.

Table 3: Sources for the trend statistics

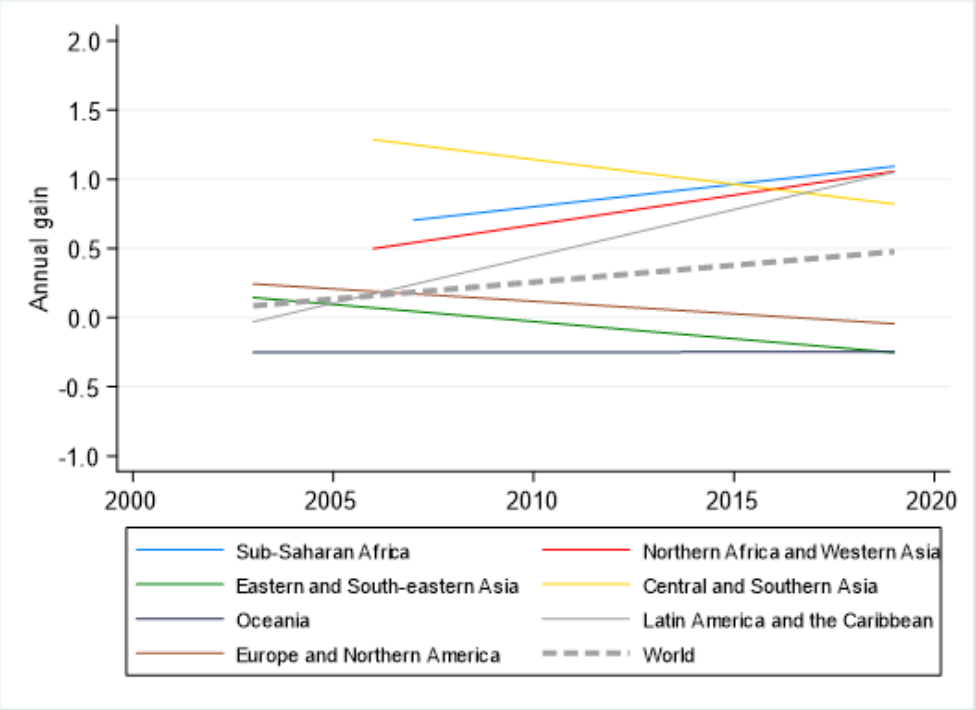
	Number of two-point trends	% of population-weighted two-point trends
LLECE	116	12.6
PASEC	44	4.5
PIRLS	100	8.3
PISA	598	54.4
SACMEQ	18	1.6
TIMSS	200	15.6
Total for international programmes	1,076	96.9
Bangladesh	4	2.1
Kenya	2	0.5
Kyrgyzstan	2	0.1
Uganda	2	0.5
Total for national programmes	10	3.1
Grand total	1,086	100.0

Of the 395 multi-point trends referred to in Table 2, 384 have consistent data sources within each trend. The difference between the two values is accounted for by the 11 two-point trends with inconsistent data sources discussed previously. If both reliability controls are applied to the 395 multi-point trends, the number is reduced to 374. For the purposes of this calculation, the slope across all points in a multi-point trend, obtained through a bivariate regression, had to exceed five in absolute terms for the multi-point trend to be excluded.

Figure 5 draws just from the 1,059 reliable two-point trends and examines through simple regression trendlines whether gains became larger or smaller over time. There is no weighting by population in this graph. Globally, gains declined slightly over time, which would be consistent with diminishing gains the higher the baseline, something seen in Figure 1. The predicted global gain values are 0.45 percentage points a year in 2003 and just 0.01 percentage points a year in 2019. Just two regions display increasing gains over time: sub-Saharan Africa and Northern Africa and Western Asia. The Latin America and the Caribbean region displays sharply declining gains. This is mostly due to comparability problems in the statistics. Details will be discussed in Section 11.

From Table 3 international programmes predominate as data sources. The last column in the table weights each two-point trend by the child population of each country. Using this approach, 97% of data sources can be said to be from international programmes. This arguably increases the reliability of trends, as large international programmes benefit from an array of skilled technical people from different countries and would be far less subject to political interference by individual national governments.

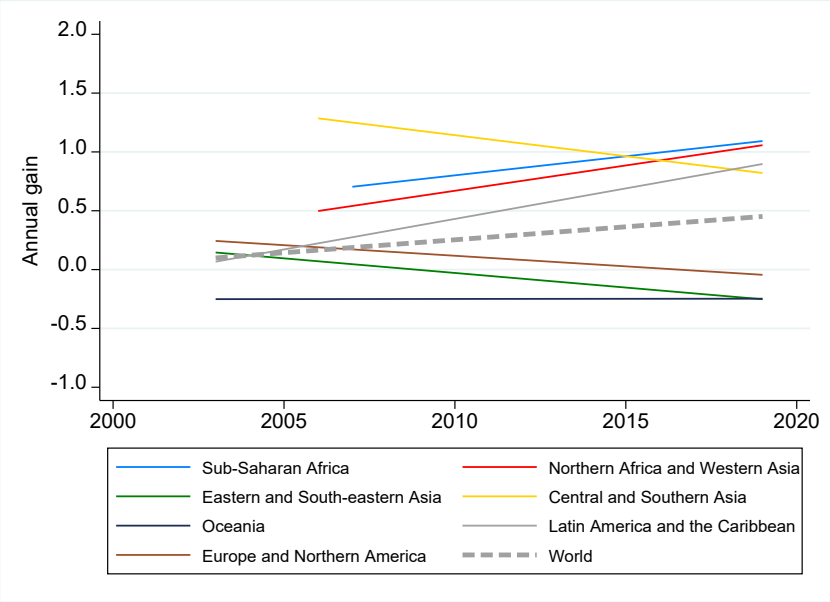
Figure 5: Changes in gains over time



Note: Initial years are not covered by the trendlines as they begin at the earliest end *point in the available two-point trends.*

If the Latin America and the Caribbean data are adjusted, using the three adjustments described in Appendix 4, the result is Figure 6 below. Here the region in question displays a relatively typical trend, which is upward. The adjustment shifts the global curve slightly. Instead of reflecting a decline in gains over time, it now reflects an increase.

Figure 6: Previous graph with LAC adjustments



4. Confidence intervals around sample-based statistics

The current report presents statistics without confidence intervals, though virtually all the testing behind the assessment programmes reflected in Table 3 draws from samples of schools and learners in each country. This means that published national proficiency statistics mostly come with margins of error, and that these margins of error can be calculated when national statistics are aggregated to the regional or global level. As explained in UIS (2020), margins of error for regional or global aggregate proficiency statistics are narrower than those at the national level. For instance, a global statistic such as that 57% of children in countries with testing are proficient comes with a 95% confidence interval of around 5.0 percentage points, essentially meaning that one can be 95% certain that the true value lies between 54.5% and 59.5%. This degree of uncertainty exists around the global annual proficiency statistics referred to in the current report.

How do confidence intervals around individual annual means impact on confidence intervals for the improvement slope? To provide an idea of this, simulations in a statistical package will reveal that given the abovementioned confidence interval of 5.0, given a period of ten years, and given an annual improvement of 0.5 percentage points a year, the slope of 0.5 percentage points will come with a confidence interval of around 0.5 percentage points, meaning one can be 95% certain that the true slope lies within the range of approximately 0.25 to 0.75 percentage points.

5. Public availability of micro- and metadata and implications for reliability

All international testing programmes display some weaknesses with respect to at least one of the three problem areas outlined in Section 3, namely sampling, data collection, and data processing (especially with regard to scaling, or the calculation of final scores). These weaknesses may be due to decisions taken at a programme level, or decisions taken within specific participating countries, possibly combined with lacking oversight from the centre. A more systematic evaluation of these weaknesses, and strategies taken in the past to address them, would be valuable. Such evaluations could assist in identifying what design improvements to prioritize in the coming years, and in interpreting trends. The fewer problems there are in the design of a testing programme, in general the greater the credibility of the data and statistics. Evaluations could also assist designers of *national* testing programmes to avoid typical pitfalls that undermine the reliability and cost-effectiveness of these programmes.

Crucially, evaluations of international testing programmes require both microdata, which should be anonymized, and metadata such as background questionnaires, non-secure test items and professionally compiled technical reports to be available, even beyond the offices of the testing programmes in question. Problems that compromise the reliability of the statistics are often identified by analysts not associated with the testing programme, and often analysts with insights into country-specific issues. Such analysts include Jerrim (2013), Carnoy *et al* (2015), Dang *et al* (2020), and Gustafsson (2020).

Table 4: Evaluation of four testing programmes

	TIMSS	PASEC	LLECE	SACMEQ
Assessment framework	2	1	2	1
Mix of MC and CR items	2	n/a	2	n/a
Item development	2	1	2	1
Translation	2	1	2	1
Piloting of instruments	2	1	2	1
Matrix design approach	2	1	1	n/a
Background questionnaires	2	0	2	0
Sampling (programme-wide details)	2	1	2	1
Sampling (country-specific details)	1	1	1	1
Weighting procedure	2	2	2	1
Data collection (programme-wide)	2	1	1	1
Data collection (country-specific)	1	0	0	0
Scaling	2	1	2	0
Microdata	2	1	2	0
Correct item responses	2	2	n/a	0
Common anchor items	2	0	1	0

The remainder of this section applies a simple evaluation framework to four testing programmes: TIMSS[22], PASEC, LLECE[23] and SACMEQ[24]. These four programmes provide around a third of the statistics used in the current report's analysis of global trends – see Table 3. The aim is in part for this to be a catalyst for more extensive evaluations in future. The approach was to consult the websites and publicly available reports of the four programmes, with a focus on a recent wave of assessment. Each cell in **Table 4** carries one of four values: 'n/a' for not applicable; 0 for not available; 1 for partly available, or if the situation seems unclear; and 2 for relatively good public availability. The evaluation is of course subject to debate, and certain details may not have been located on the websites. However, the evaluation is informed by discussions which are common among analysts both within the testing programmes, and analysts located outside of them. In many cases, a specific cell in the table has already been subject to some form of evaluation, possibly within the programme itself.

The longstanding and relatively well-resourced TIMSS testing programme scores highly against nearly all aspects of the evaluation. TIMSS procedures which apply across all countries, and many country-specific details, are extensively documented in the 1172-page *Methods and procedures* report[25], which in many ways serves as a standard-setter for other programmes. Separate TIMSS reports provide further details, for instance, the background questionnaires used. To illustrate, the programme-wide rules for sampling are spelt out in *Methods and procedures*, and in addition details per country are provided, such as schools excluded, stratification details, and the extent to which replacement schools were used. Many participating countries produce their own national reports where further details on the sample are provided. However, some country-specific details seem less easy to obtain, for instance exactly when in the year learners were tested and whether service providers implementing the testing changed from one cycle to the next. Such details can be important for understanding the reliability of trends[26]. With regard to data collection, what is often not easy to obtain is reasons why response rates to the background questionnaires are relatively low, in particular for certain countries.

[22] Trends in International Mathematics and Science Study.
[23] Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for Assessment of the Quality of Education).
[24] Southern and Eastern Africa Consortium for Monitoring Educational Quality.
[25] Martin *et al*, 2020.
[26] Jerrim, 2013.

Importantly, it appears that enough information is provided in *Methods and procedures* for an appropriately skilled analyst to replicate, via coding, the conversion of raw item-level responses in the data to the 'plausible values' which can be considered the final scores of learners[27]. Alternatively, a more straightforward replication is possible given the fact that certain statistical packages provide the tools to replicate the TIMSS scaling approach.

In the case of PASEC 2014 there is a 184-page manual[28] for data analysts available online, though it is not easy to find[29]. In the case of PASEC 2019, there is an 18-page methodology section within the easily located larger report[30] containing PASEC results and trends. It appears that no data analyst's manual for 2019 was available online by early 2023. A manual for analysts will generally provide some background information on the methodology employed, though this would tend to be limited. Within the two texts that were found, the following issues listed in Table 4 are covered: basic details relating to the assessment framework and item development; the basic structure of the matrix design (how certain items are selected for certain test booklets); a brief statement on the piloting of test items; basic programme-wide sampling standards, for instance the use of probability proportional to size; limited country-specific details relating to sampling, for instance strata used; considerable detail relating to weights; a broad description of how the data collection process was overseen and quality assured across the programme as a whole; the basics of the generation of plausible values in the scaling process[31]; and which choices are correct in the multiple-choice questions.

Because PASEC employs only multiple choice (MC) questions, no account is needed of the balancing of these questions and constructed response (CR) questions. Background questionnaires employed could not be found, meaning the analyst would need to rely on abbreviated questions in the microdata. The PASEC website has provided details on how researchers can access the microdata, but access to the data has reportedly not been easy, with few researchers outside PASEC having gained access. In contrast, TIMSS data are easily downloaded off the TIMSS website.

For those who have accessed PASEC microdata, what is not easily established is what common anchor items span both 2014 and 2019 to permit the gauging of trends. Ideally, these details should be included in a methodology report.

In the case of LLECE, texts that emerged as interesting include a 31-page series of annexes in the report on the 2019 testing cycle[32], as well as a 422-page technical report[33] covering the previous cycle, of 2013. These two texts provide more information on what is being assessed and the item development process than what appears in PASEC, yet less than what appears in TIMSS. Questionnaires (for 2013) are easily accessible online, and considerable detail is provided in the 2013 technical report regarding the sampling approach. The scaling process is described in considerable detail. The 2019 microdata with some supporting documentation are easily downloadable. SERCE appears to follow an approach with respect to multiple-choice questions of simply indicating whether learners responded correctly or incorrectly to each item. This removes the need for a list of correct

[27] However, replicability through fresh coding was not tested.
[28] CONFEMEN, 2017.
[29] Internet searches referred to here all occurred in January 2023.
[30] CONFEMEN, 2020.
[31] While the 2014 PASEC manual recommends that analysts use specific Stata .ado files to take into account the way scaling occurs, these .ado files do not appear to be available online.
[32] UNESCO, 2021.
[33] UNESCO, 2016.

item responses. However, the disadvantage with this approach is that analysis of patterns of cheating in the test administration process becomes less feasible. To illustrate, knowing that all learners in a school responded incorrectly to a specific item is not as informative as knowing that all these learners selected the *same* incorrect option.

Finally, the reliability of statistics emerging from SACMEQ are particularly difficult to verify: the microdata are particularly difficult to access, and technical documentation on methodologies used is lacking. The approach within SACMEQ has been to consider each national education authority the owner of its data. It is thus difficult for any analyst to compile a full dataset as permissions would be required from some fifteen different organizations. A combined dataset is necessary for comparison purposes, and certain data quality checks. The central SACMEQ office has the data for all countries, and the programme's first international report was released in 2021, using the data from the last SACMEQ run of 2013[34]. This 137-page report has a 9-page summary of the methodology employed. The last time comprehensive methodology reports were produced was following the 2000 run of the programme – this includes an important and publicly available 19-page account of the methodology pursued 20 years ago[35]. It can be assumed that many practices would have been carried through to the subsequent 2007, 2013 and 2021 runs, but the details are not readily available. Above all, how scoring occurred to make statistics comparable over time is not clear. The SACMEQ column of Table 4 largely reflects what is seen in the two texts referenced here. As SACMEQ provides the same test to all learners, there is no matrix approach.

There is room for improvement in the design of the four programmes discussed in this section, though the importance of this varies. PASEC and especially SACMEQ are most in need of strengthening. In part, this should be realized through learning across programmes, and careful consideration of costs relative to benefits.

6. Characteristics of countries with greater 4.1.1 data availability

Given that around half of the world's children are in countries where no official SDG proficiency trends are available (see details in Section 11), a key question is whether this half of the world's children are in countries with a lower likelihood of proficiency improvements, compared to the half who are in countries with proficiency trend statistics. Put differently, how representative is the half of the world *with* trend data of the entire world?

To answer this question, a simple measure of trend availability was constructed. The measure used each country's longest, in terms of number of data points, level- and subject-specific trend as an indicator. If a country had no SDG 4.1.1 values at all, zero was assigned to the country, while 1 means some 4.1.1 values exist but no trend is available, 2 means there is at least one two-point trend, 3 means the longest trend has three points, and so on. Any number of points above 5 was converted to 5 as the aim was to explore differentiation among more 'data poor' countries. In constructing the measure, only the 374 trends found after the reliability controls were considered (see Section 3) were used.

Table 1 refers to five possible predictors of the availability of proficiency data. Country size is represented by the population aged 0 to 14 (the source here is UIS.Stat – see Section 8). It is possible that very small countries lack the capacity or financing to participate in the relevant international

[34] Awich, 2021.
[35] Ross *et al*, 2008.

testing programmes. The most recent pre-2020 measure of the percentage of children of primary school age who are in school was used an indicator of educational development (the source here is also UIS.Stat – see discussion in Section 10)[36]. Countries with low levels of participation could be under-prioritizing education and may therefore not be interested in monitoring proficiency. The average annual improvement between around 2010 and 2021 in this indicator was also calculated. The 2019 income per capita with purchasing power parity, as reflected in the World Development Indicators[37], was used as a measure of a country's general level of development. Finally, of the six main indicators within the Worldwide Governance Indicators[38] programme, the indicator 'government effectiveness' was considered the most likely predictor of a country's preparedness to monitor SDG 4.1.1. For all countries, the governance situation in 2019 was used. The governance indicator carries values ranging from -2.5 to 2.5, with greater values representing more government effectiveness.

Table 5: Means and standard deviations of variables used in the regression

Variable	Mean	Std. dev.	Countries with non-missing values
Points in trend	2.15	2.00	228
Child population (thous.)	8758	31020	228
% in school	92.52	10.70	190
Annual change in above	0.00	1.35	178
PPP income per capita	22153	22765	192
Government effectiveness	-0.03	0.99	206

It should be noted that although the annual change in the percentage of children in school ('Annual change in above') carries a mean of zero, there were somewhat more countries with an upward trend than a downward trend – 98 against 78 respectively.

The outputs of four regression analyses are shown in **Table 6.** In Model A, which region a country is in emerges as a strong predictor of the 'data-richest' series found. The coefficients on the regional dummies, Europe and Northern America being the omitted reference region, are all negative and nearly all statistically significant. Among the remaining potential predictors, just two are statistically significant reaching at least the 10% level: higher participation in primary schools and greater government effectiveness are associated with better trend data. Model B allows for more countries to be covered through the exclusion of variables with relatively high degrees of missing values. Model C excludes all countries from Europe and Northern America in order to bring about a stronger focus on developing countries. Finally, Model D excludes just those countries with fewer than 100,000 children aged 0 to 14, in other words particularly small countries with little influence on population-weighted aggregate statistics. Across all four models government effectiveness emerges as a statistically significant predictor. If income per capita is replaced by its natural logarithm, to compress differences among rich countries, income per capita remains statistically insignificant and government effectiveness remains a significant predictor.

[36] As will be seen in Section 10, completion of primary is the optimal indicator in terms of coverage of the world's children is concerned. However, indicators of being in school are optimal in terms of maximising the number of countries analysed.
[37] https://databank.worldbank.org/source/world-development-indicators
[38] https://databank.worldbank.org/source/worldwide-governance-indicators

Table 6: Regressions on points in trend

	A	B	C	D
Dependent variable → Points in trend				
Constant	1.219	3.679***	-1.730	0.390
Child population (thous.)	0.000	0.000	0.000	0.000
% in school	0.025*		0.028*	0.032**
Annual change in above	0.080		-0.048	0.012
PPP income per capita	0.000		0.000	0.000
Government effectiveness	0.589**	0.676***	0.467*	0.662***
Sub-Saharan Africa	-1.572***	-1.793***	1.046**	-1.207***
Central and Southern Asia	-2.195***	-1.895***	0.375	-1.976***
Eastern and South-eastern Asia	-1.589***	-1.544***	0.851	-1.488***
Latin America and the Caribbean	-1.712***	-1.933***	0.804*	-0.541
Oceania	-2.567***	-2.808***		-1.848***
Northern Africa and Western Asia	-0.023	0.017	2.465***	-0.019
N	166	206	124	141
Adjusted R squared	0.512	0.450	0.343	0.620

*Note: *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level, and * at the 10% level. Observations (countries) are unweighted.*

It can be concluded that countries without SDG 4.1.1 trend statistics are unlike those with such statistics in important ways. The former tend to display lower levels of government effectiveness, even after controlling for factors such as country size and per capita income.

Table 7 indicates that without any filtering of countries, across all regions other than Latin America and the Caribbean, having SGD 4.1.1 trend statistics with two or more points is associated with more government effectiveness. Latin America and the Caribbean also fits this pattern if very small countries are excluded (in the case of this region this would largely be Caribbean islands). Thus, even within regions countries with proficiency trends are unlikely to be representative of all countries in the region. Insofar as government effectiveness strengthens the likelihood that improvements in proficiency will be seen, countries with trend data are likely to display better progress than countries without trend data.

Table 7: Government effectiveness within regions

	No filters (206 countries)		Only where child population exceeds 100,000 (171 countries)	
	No trends	With trends	No trends	With trends
Sub-Saharan Africa	-0.970	-0.594	-0.985	-0.639
Northern Africa and Western Asia	-1.349	-0.008	-1.349	-0.008
Eastern and South-eastern Asia	-0.151	1.011	-0.315	0.988
Central and Southern Asia	-0.573	-0.462	-0.573	-0.462
Oceania	-0.296	1.621	-0.537	1.621
Latin America and the Caribbean	0.076	-0.112	-0.652	-0.112
Europe and Northern America	0.369	0.967	-0.409	0.913
World	-0.403	0.262	-0.741	0.209

7. Concentrations of populations with specific countries

It is useful to assess the concentration of children within specific countries per region, as this provides an idea of how many countries, in particular smaller countries, can remain with no statistics while credible regional aggregates can still be calculated. **Figure 7** illustrates the global situation. Just two

countries, India and China, account for 31% of the world's children and 10% of the largest countries account for 70% of the world's children.

Figure 7: Global concentration of populations in countries

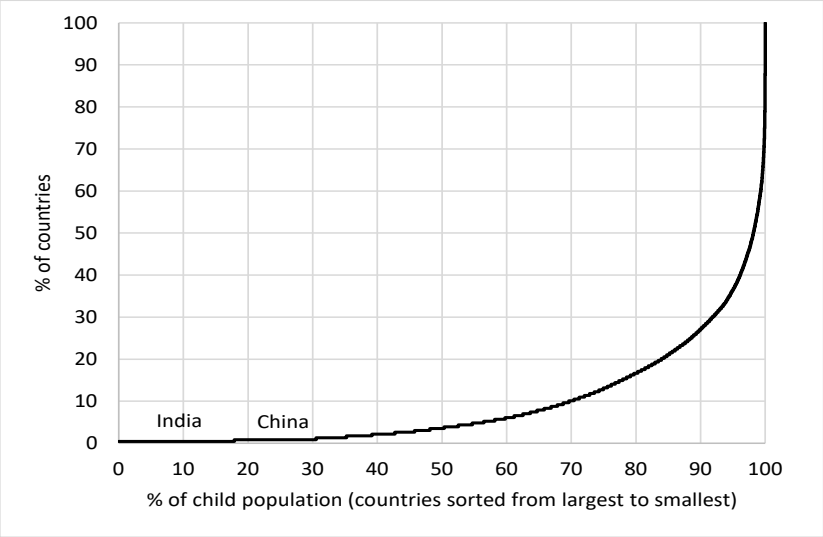
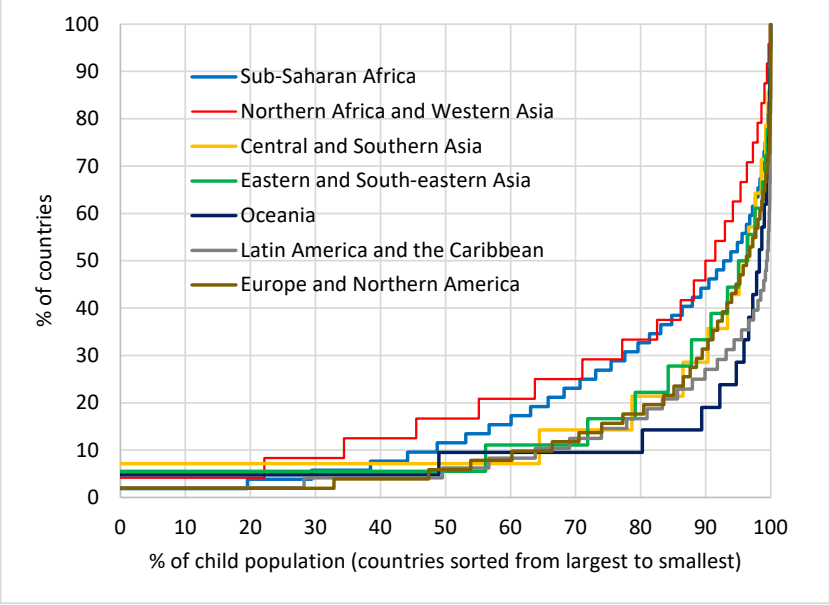


Figure 8 produces similar curves for regions. The region where children are most spread out across countries is Northern Africa and Western Asia. At the other end of the spectrum, child populations are highly concentrated in just a few countries in Oceania.

Figure 8: Regional concentration of populations in countries



8. Trends with respect to child populations

Indicator 4.1.1 as currently used expresses the percentage of *enrolled learners* who reach minimum proficiency levels, not the percentage of the child population. This practice has come about more due to the way data are collected than to any deliberate decision around definitions. The wording of

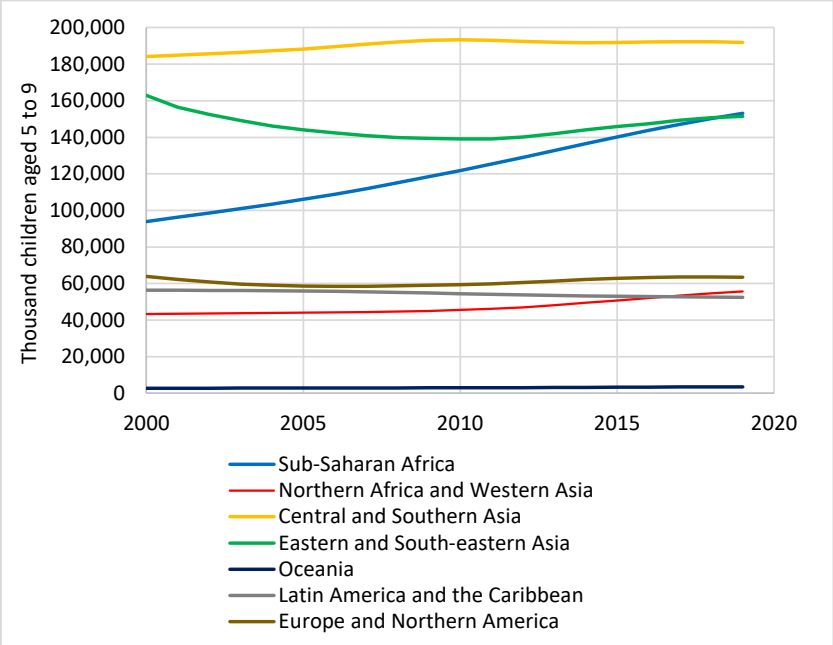
Indicator 4.1.1 is arguably ambiguous when it comes to the question of whether attending learners of an age cohort of children should be the denominator.

Given the way the indicator is calculated, it would have been consistent to weight country statistics using *enrolments* at specific levels. However, a good enrolment dataset, by grade, covering all countries in the period 2000 to 2019 was not found. Instead, historical population totals in five-year age bands published as part of the World Population Prospects 2022[39] was used. This is likely to approximate well what would have been obtained had weighting by enrolment been used. In particular, the major effect of a rising number of children in sub-Saharan Africa is likely to be captured as well using population statistics as using enrolment statistics.

The historical World Population Prospects data are used only for the calculation of aggregate proficiency trends. In other situations where population is used to weight countries in the report, the most recent age 0 to 14 total per country available in UIS.Stat in 2022 was used. The most recent year per country was mostly 2022.

As can be seen in **Figure 9**, children aged 5 to 9 increased in sub-Saharan Africa between 2000 and 2019 by over 50%, an increase that clearly exceeds that of the other regions.

Figure 9: Trend for children aged 5 to 9



Population of ages 5 to 9 was used for weighting Grades 2/3, ages 10 to 14 were used for end of primary, and 15 to 19 for lower secondary. At the secondary level, where the population out of school is higher than at the primary level, is likely to produce *levels* of proficient youths which is over-stated if population instead of enrolments are used. However, the distortion would be less serious with regard to *trends*, assuming a relatively constant secondary school participation rate. Clearly this is one

[39] https://population.un.org/wpp

area where there is room for improvement in future estimation of regional trends, assuming that sufficiently comprehensive and reliable enrolment numbers, or participation rates, are obtained.

9. A methodology for aggregating levels and trends

Not much work has occurred to gauge pre-pandemic learning proficiency trends at the regional and global levels. UIS (2021b: 44), a report focusing on the impacts of the pandemic on learning, assumes that prior to the pandemic global learning proficiency was improving by around 0.7 percentage points a year. This is based in part on trends among developing countries seen in PIRLS[40], and would be in line with work by the World Bank[41]. Improvements of around 0.7 percentage points a year are also seen in rough (and completely linear) global estimates for all six 4.1.1 indicators available in the SDG section of UIS.Stat[42]. What appears in the current report is the most comprehensive attempt to date to gauge regional and global pre-pandemic trends. Despite this, even the work presented here suffers from several limitations, such as serious gaps in the source data. The details of this are discussed extensively.

Figure 10 illustrates one of two methods which will be followed in the current report to calculate aggregate trends for groups of countries. In both methods, only countries with trends based on consistent sources are used to feed into the aggregate trend statistics.

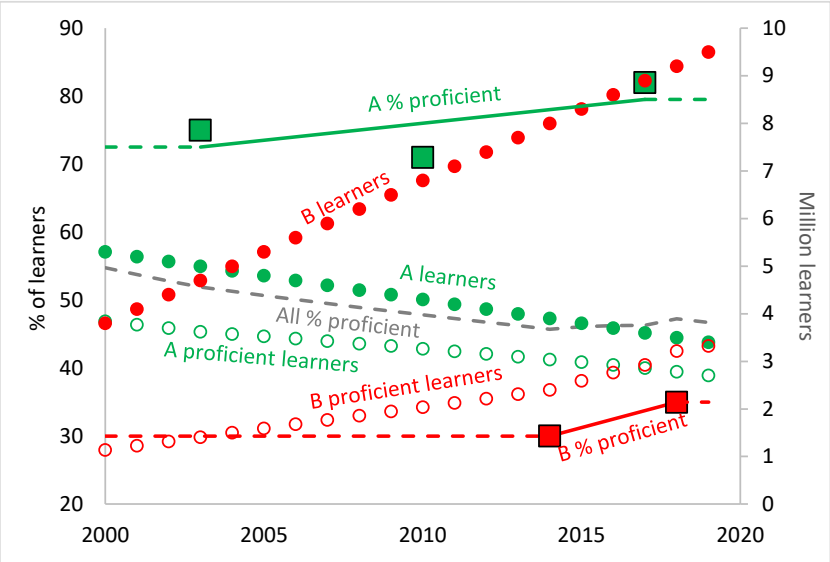
Figure 9 illustrates two hypothetical countries, A and B, the only two countries in the group for which an aggregate trend must be estimated. Country A has proficiency statistics for one level and subject in three years: 75% in 2003; 71% in 2010; and 82% in 2017. These three points are indicated by square green markers. A simple linear trend, obtained by regressing year on proficiency, reflects an upward trend. Before 2003 and after 2017 the trend is assumed to be flat, and at the two levels of the ends of the trendline. The flat initial and final trends are what make this the 'restricted' trend method. The curve 'A learners' indicates that the learner population relevant to the level is in decline, from 5.3 million in 2000 to 3.4 million in 2019 – this trend should be read against the right-hand axis. As discussed in Section 8, age-specific population statistics will be used to represent learner populations. The 'A proficient learners' is the population for each year multiplied by the proficiency statistic for each year, the proficiency statistic being read off the 'A % proficient' curve.

[40] UIS, 2019.

[41] World Bank, 2019: 18. Note, however, that World Bank (2022: 9) concludes that learning proficiency has declined, even before the pandemic. Specifically, a decline of around 1.0 percentage point a year globally for the period 2015 to 2019 is found.

[42] Accessed November 2022. Regional trends are also linear and largely align to estimates in UIS (2021).

Figure 10: Aggregation method with restricted trend

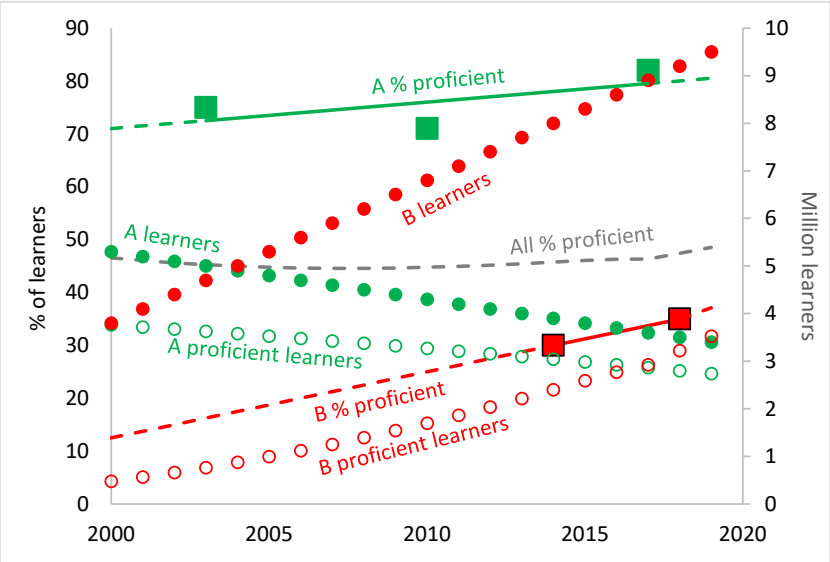


Country B, with a larger and increasing learner population, and low but improving proficiency levels, is handled in the same way as Country A. Because Country B only has two proficiency statistics, 30% in 2014 and 35% in 2018, the simple connecting line represents the trend. As in the case of Country A, Country B initial and final trends, for which no proficiency statistics exist, are assumed to be flat.

It is clear that although Country A's trend for its proficiency statistics is an upward one, the trend for the *number* of proficient children is downward, because the child population is declining. Adding the proficient learners across A and B and dividing by the total learners across A and B, for each year, produces the aggregate trend 'All % proficient'. In this example, that trend is *negative*, despite the fact that both countries display positive trends for their proficiency statistics. This is because of changing learner population numbers. As population growth is strong in a country with a generally low level of proficiency, Country B, and as population is declining in a country with a generally high level of proficiency, Country A, the overall effect is one of a decline in the overall percentage of learners who are proficient. As will be seen in Section 10, demography does not influence trends to the extent illustrated in this example, yet to some extent the effects illustrated in the example play a role.

Figure 11 illustrates the second aggregation approach, one which uses a 'continuous trend'. The only difference compared to the previous approach is that here initial and final trends follow the slope for the period where proficiency statistics do exist. Put differently, each country has a completely linear trend from 2000 to 2019. In this example, the aggregate trend is roughly a flat one. The clear decline in 'All % proficient' seen in Figure 9 is not visible in Figure 10.

Figure 7: Aggregation method with continuous trend



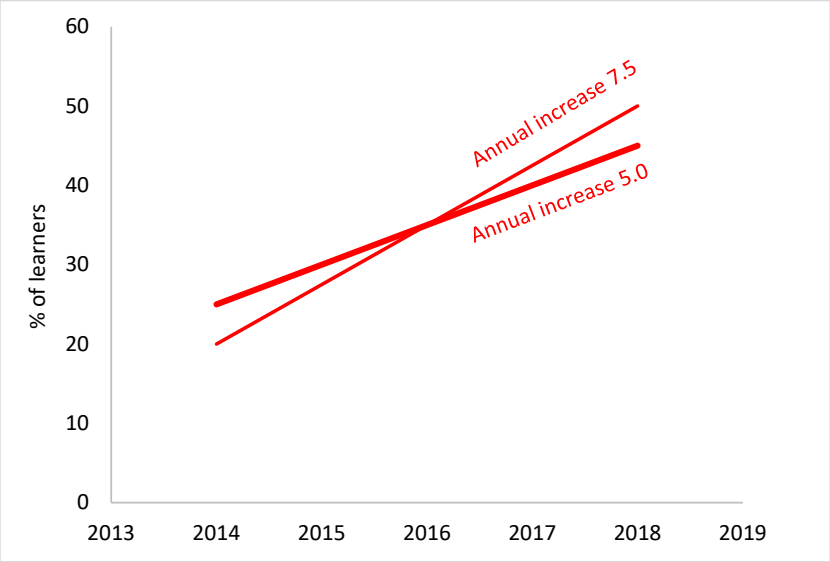
Which of the two approaches is likely to be most reliable? The answer to this question depends strongly on whether the timing of the monitoring of proficiency within countries is random or not. If the years when proficiency was measured is fairly random, then the continuous approach can be considered most reliable. To illustrate, there could be an international testing programme that tested in 2005, 2011 and 2017. There could moreover be a country that participated in just two consecutive assessments, either 2005 and 2011 or 2011 and 2017. If the timing of the country's assessments were not related to the likelihood of there being a proficiency improvement, it would not matter for the purposes of gauging trends when the country measured proficiency. However, it is very possible that countries will pay more attention to monitoring proficiency when there is simultaneously a strong focus on improving learning. This would make the timing of measurement far more non-random. Despite these reservations, the discussion that follows will consider slopes seen in the continuous approach as a somewhat more reliable reflection of reality than the restricted approach. The restricted approach has its own limitations. If a country monitored proficiency in, for instance, 2010 and 2015, and saw an improvement in this period, the assumption that outside this period there would be no change at all is unrealistic. The trends seen from the two approaches can be thought of as upper and lower bounds of what the true slope is.

A key feature of the UIS.Stat proficiency dataset used is that in the 11 cases where the data source is inconsistent over time (see Section 3), *the trend always consists of just two points*. This means that in aggregating trends, and if data source inconsistencies are to be excluded, no decisions need to be made around which point to exclude. Both points are deemed not able to produce trends. Had there been situations where, for instance, data sources X, Y, X and Y had been used sequentially across four different years, a decision would have to be taken around whether to calculate a trend on the basis of source X or source Y. As it turns out, such situations do not arise.

There are 384 multi-point trends available for the various aggregations, after 11 cases with inconsistent source data have been removed (see Section 12). Of the 384, ten have gains or losses which exceed five percentage points a year. **Figure 12** illustrates what occurs in these cases. A trend with an annual increase of 7.5 a year would be made less steep, to the point at which the slope becomes 5.0, by pivoting the trendline on its midpoint. The assumption is thus made that, for instance,

an annual increase of 7.5 probably represents a mismeasurement, but that a considerable increase did in fact occur, though it is not as large as the reported increase.

Figure 8: Attenuating extreme changes



In the case of both methods, any annual proficiency statistic emerging as less than zero in the calculation, is adjusted to zero. Similarly, a ceiling of 100% applies.

10. Graphs with national, regional and global trends

The following eight composite graphs (**Figures 13** to 20) represent the seven world regions, and the world. The objective of these graphs is to provide a detailed visualization of the Indicator 4.1.1 statistics. They include a few explicit statistics. These statistics, plus additional statistics, are presented in Section 11 in tables.

Guidance to reading the composite graphs is as follows:

- The points in the graphs are countries with proficiency statistics, their diameter being proportional to the population aged 0 to 14 (the most recent statistic was used, meaning diameters do not change for a country over time). Grey points represent countries with no second point, meaning no trend is discernible.

- Lines represent *two-point* trends. The very few dotted lines would represent trends where different data sources feed into the two endpoints. Red lines, which are more common, represent trends which are steeper than would typically be considered credible – any annual change exceeding 5.0 percentage points, either positive or negative, is considered possibly problematic. Lines are coloured red where the slope for an entire series, which could consist of more than two points, and hence more than two line segments, exceeds a steepness of absolute 5.0. Red lines reflect the situation before the adjustments discussed in Section 9 are applied. Red and dotted lines would only appear in the first six graphs of each figure relating to Indicator 4.1.1.

- Green lines reflect aggregate trends. In the case of the first six graphs of each figure, light green lines represent the 'continuous' method described in Section 9, while dark green lines represent the 'restricted' approach. It is clear that the two approaches do not produce very different patterns in most graphs. Thin green lines represent the aggregate trend as explained in Section 9, and these may be non-linear. Thick green lines represent the linear trend based on the thin green lines. In most instances thick lines obscure thin lines because underlying aggregate curves are linear, or almost exactly linear.

- The three statistics per graph in red text begin with a percentage of children with any point statistic. The second statistic is the percentage of children in countries with a usable trend which could be used for the aggregation. In other words, trends other than those with inconsistent sources apply. The third statistic is a population-weighted measure of the availability of trends over the years covered by the data. This range of years is mostly shorter than the full 2000 to 2019 scope. To illustrate, if the earliest statistics for end of primary reading in a region are from 2003, and the latest from 2018, then that range of 15 years is considered. If in Country A there are statistics for 2003 and 2015, that means Country A has a trend available for 12 years, or 80% of the range. If in Country B the earliest and most recent statistics are for 2015 and 2019, just four years, or 27% of the range, are covered. The population-weighted mean of the 80% and 27% would be found. Because different graphs within the figures have statistics spanning different ranges of years, the indicator values are indicative rather than strictly comparable.

- The participation statistics illustrated in the last two graphs of each figure all extend back to 2010 only. That is the range within the original UIS.Stat data currently accessible. The names of the two indicators within UIS.Stat are 'Out-of-school rate for children of primary school age' and 'Completion rate, primary education' (the latter of these two should be distinguished from the VIEW completion statistics discussed in Section 13). The two indicators are packaged as part of SDG indicators 4.1.4 and 4.1.2 respectively. For the first of these two indicators, the difference between the indicator value and 100% was found, which would be the percentage of children who are *not* out of school. Only one kind of trendline is included in the participation graphs, a linear trendline in dark green. The slope of this trendline is the population- and duration-weighted mean of the various two-point slopes found in the data. The two participation graphs are included for two reasons. Firstly, they illustrate how low the availability of trend statistics is in the case of learning proficiency, relative to participation. Secondly, they point to improvements which are often steeper than the improvements seen with respect to proficiency, keeping in mind that the proficiency statistics refer to children and youths who are *in* school.

Figure 20 depicts the global situation. What are the striking features? Globally, the availability of proficiency statistics improves the higher the schooling level. This is despite the fact that monitoring learning in the initial grades is widely seen as especially important, as this is where the foundations for educational success are laid. (The 'trend coverage' statistics are easiest to see in **Table 9**.) The green aggregate curves in Figure 19 do not point to the strong upward slopes that would be necessary if the SDG targets of 100% by 2030 were to be achieved. The trends are relatively flat. At the same time, the two participation graphs in Figure 19 display clear upward trends. This suggests that the percentage of *children in the population* who are proficient is likely to be increasing. Figure 19 allows for rough estimations of the magnitudes of these changes. Completion of primary schooling improved from 75% to 89% between 2010 and 2019 – this is what the aggregate curve in the completion graph in Figure 19 points to. Over this 2010 to 2019 period, reading proficiency at the end of primary

remained at around 60%. Assuming that the *additional* children who completed primary schooling were from the more socio-economically disadvantaged segments of society, it can be concluded that the percentage of proficient children in the population improved from roughly 45% (60% × 75%) to 53% (60% × 89%), a substantial gain of eight percentage points in the population, almost a percentage point per year. This is explored further in Section 13.

Figure 13 Trends in sub-Saharan Africa

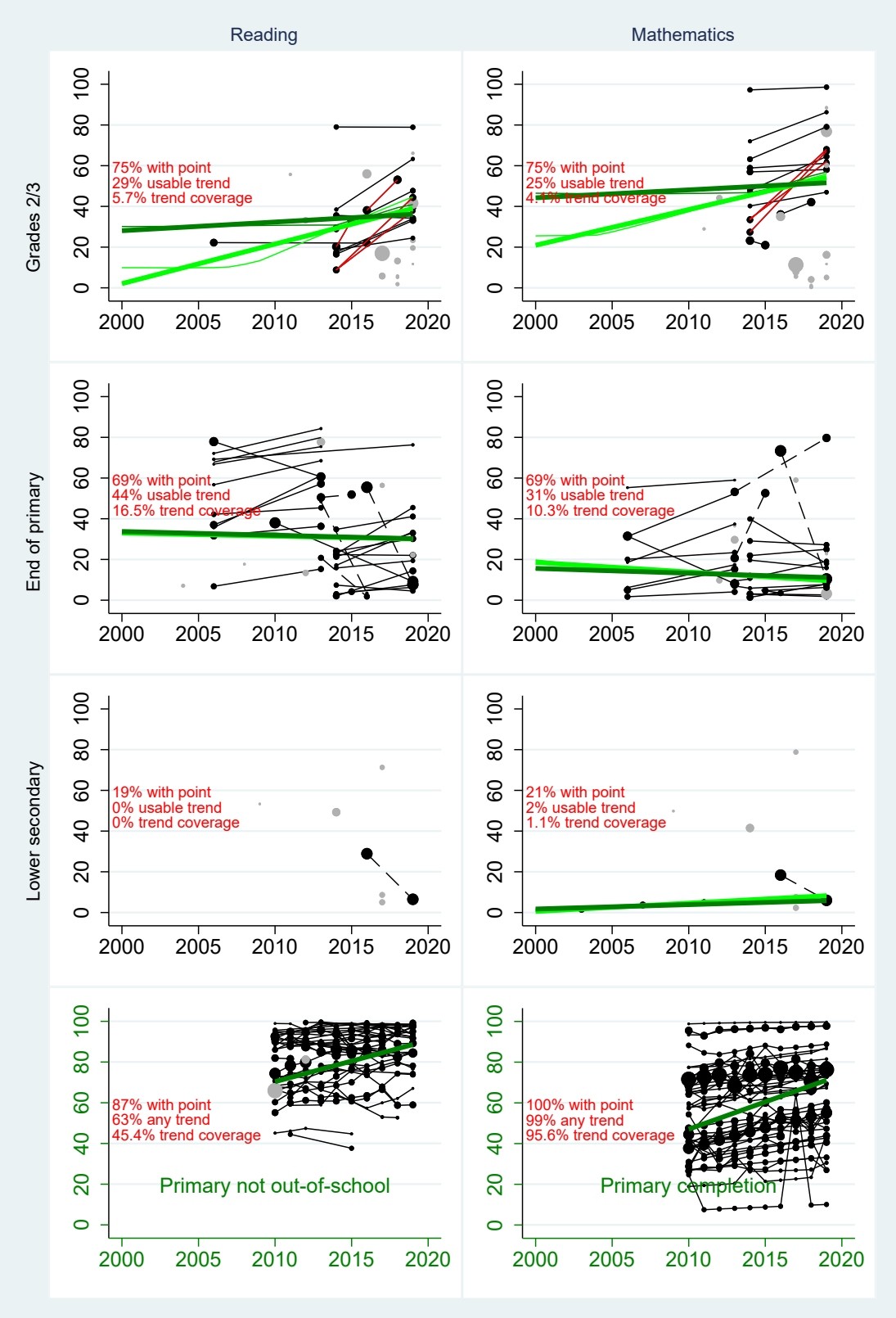


Figure 9: Trends in Northern Africa and Western Asia

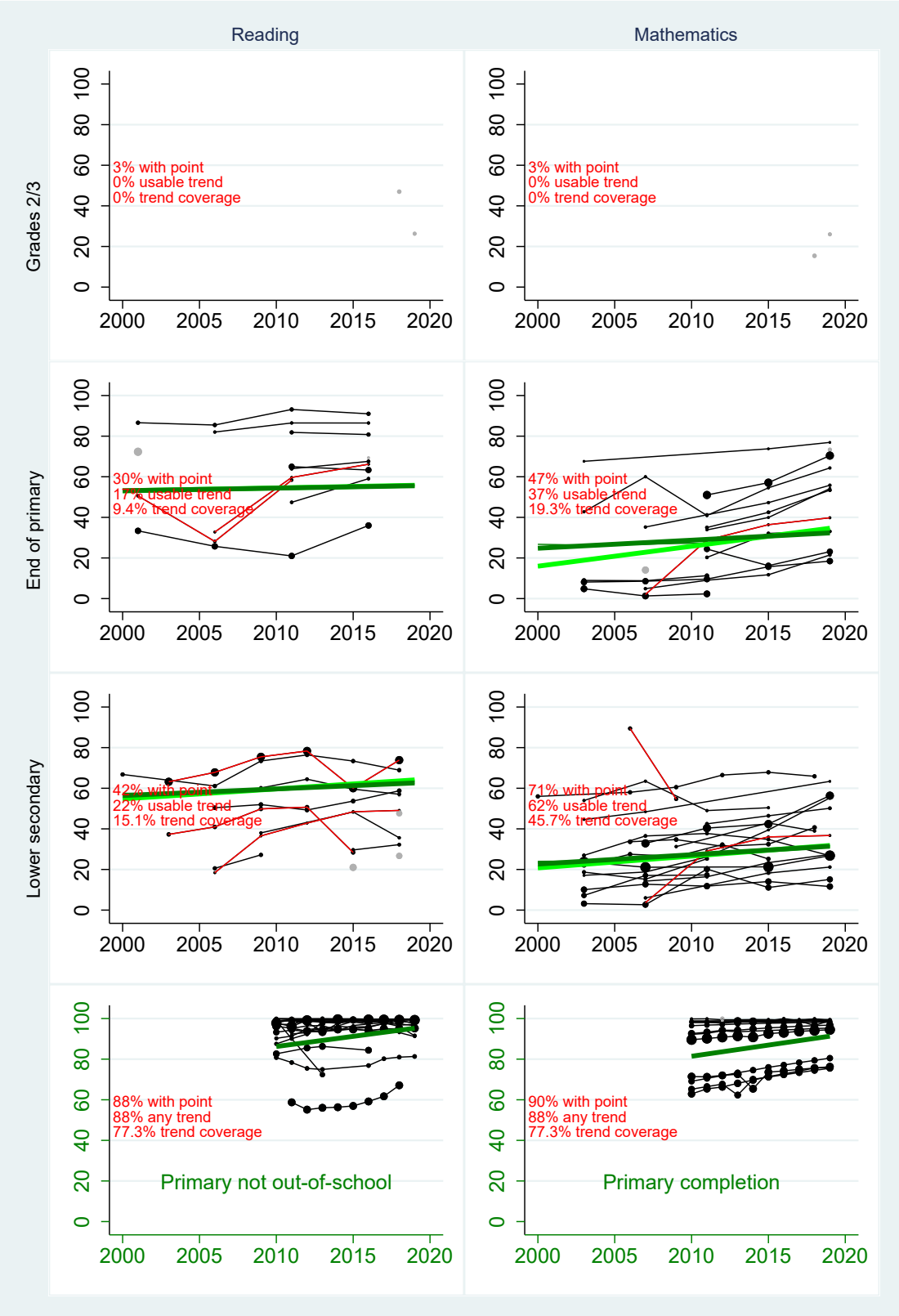


Figure 10: Trends in Central and Southern Asia

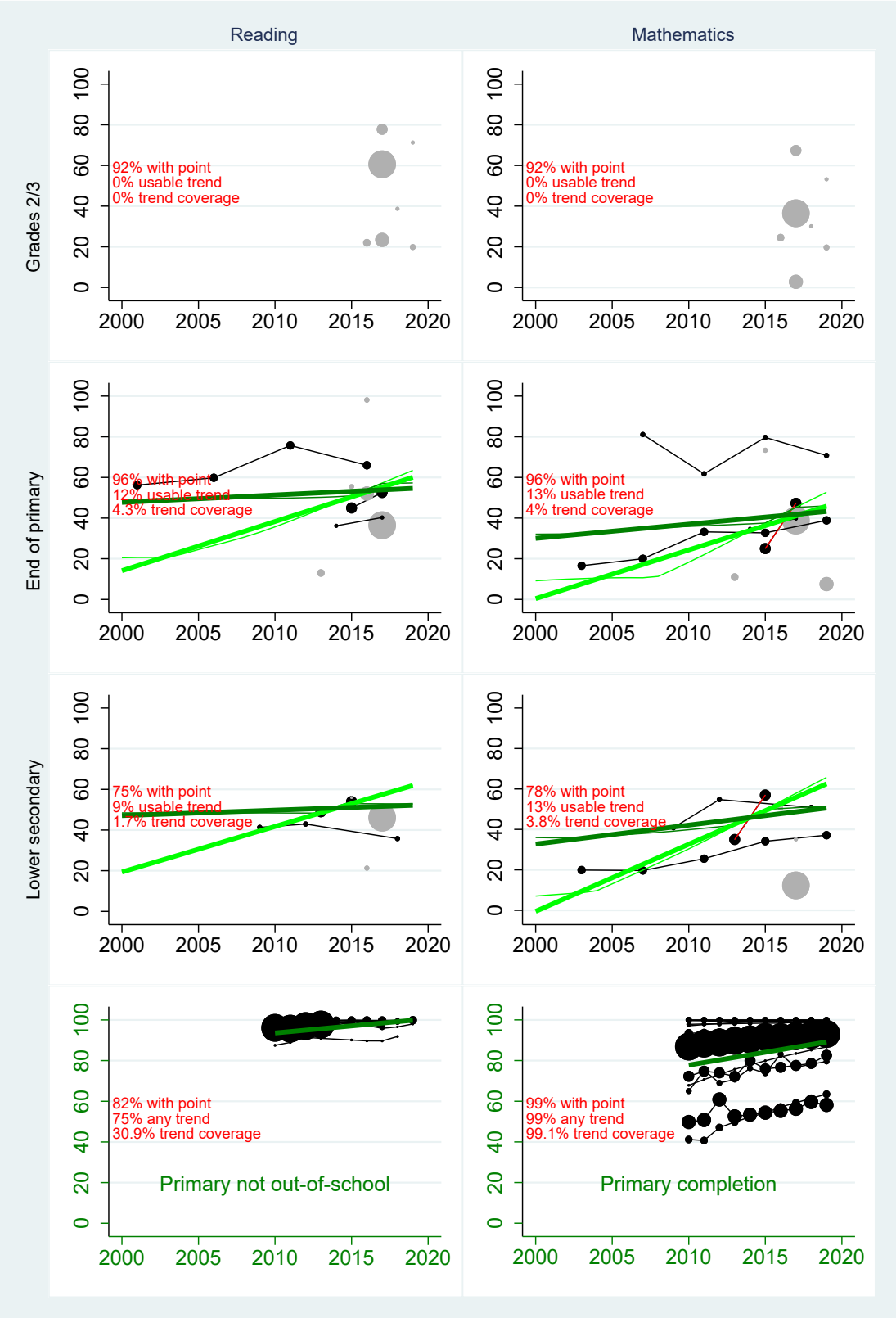


Figure 11: Trends in Eastern and South-eastern Asia

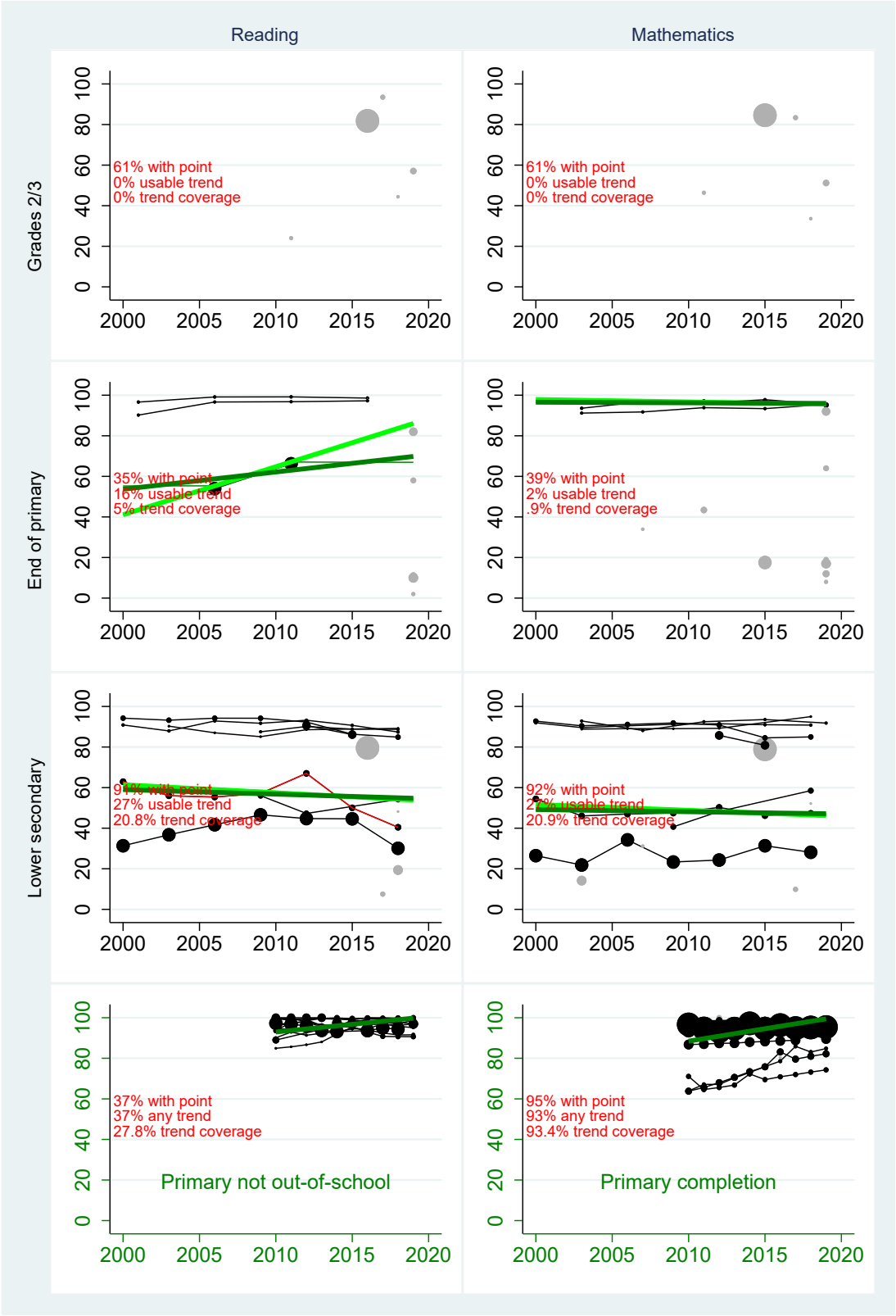


Figure 17: Trends in Oceania

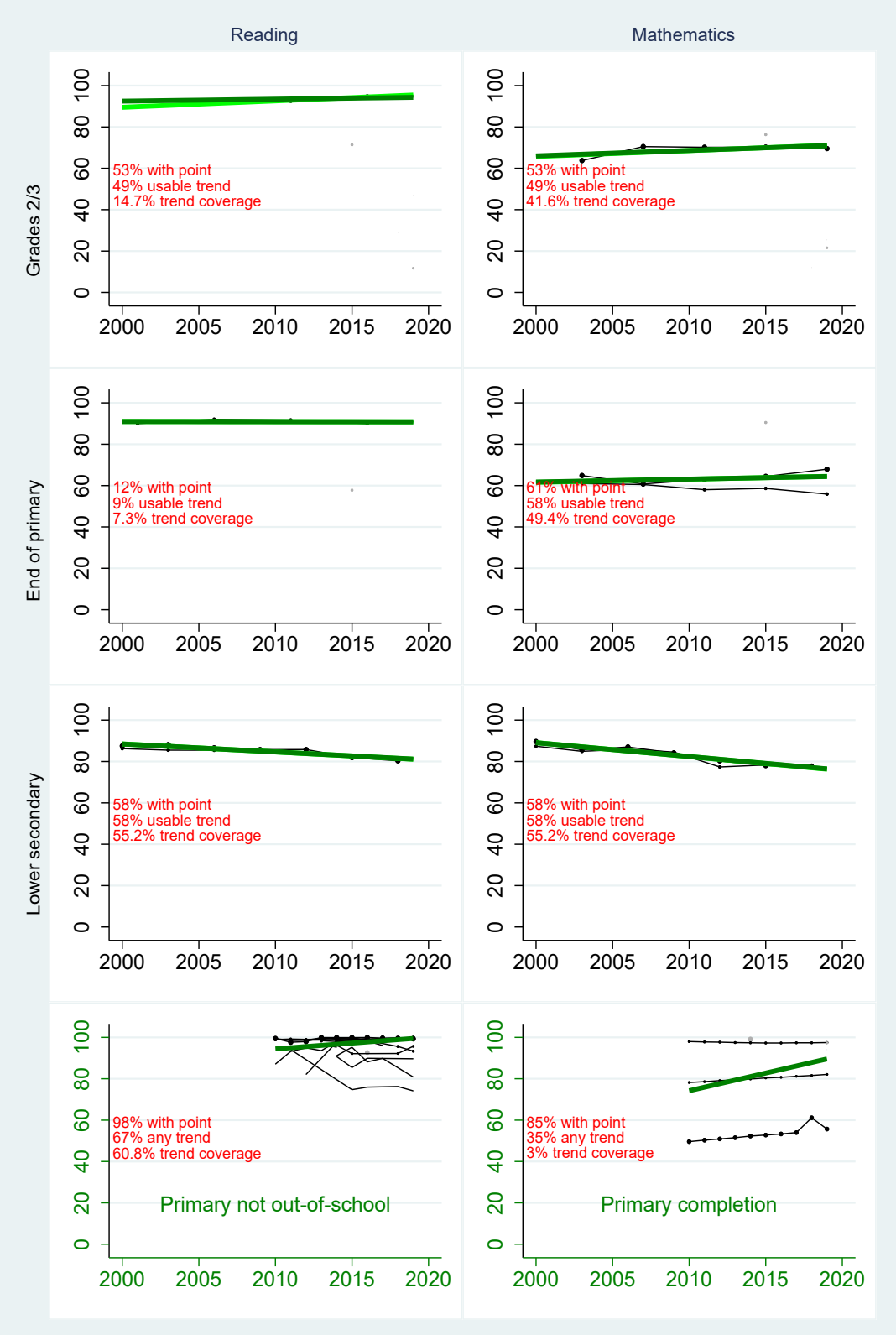


Figure 18: Trends in Latin America and the Caribbean

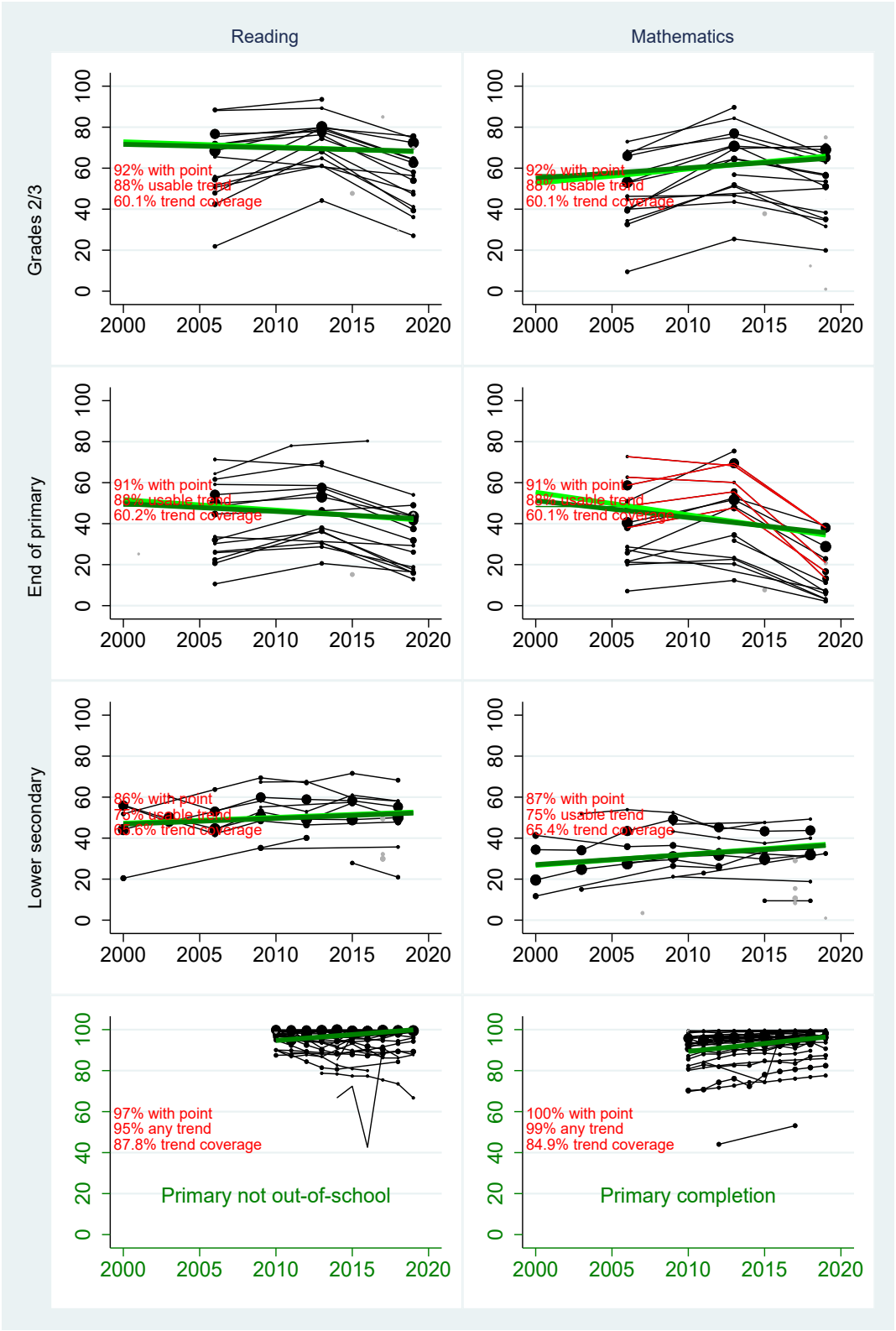


Figure 12: Trends in Europe and Northern America

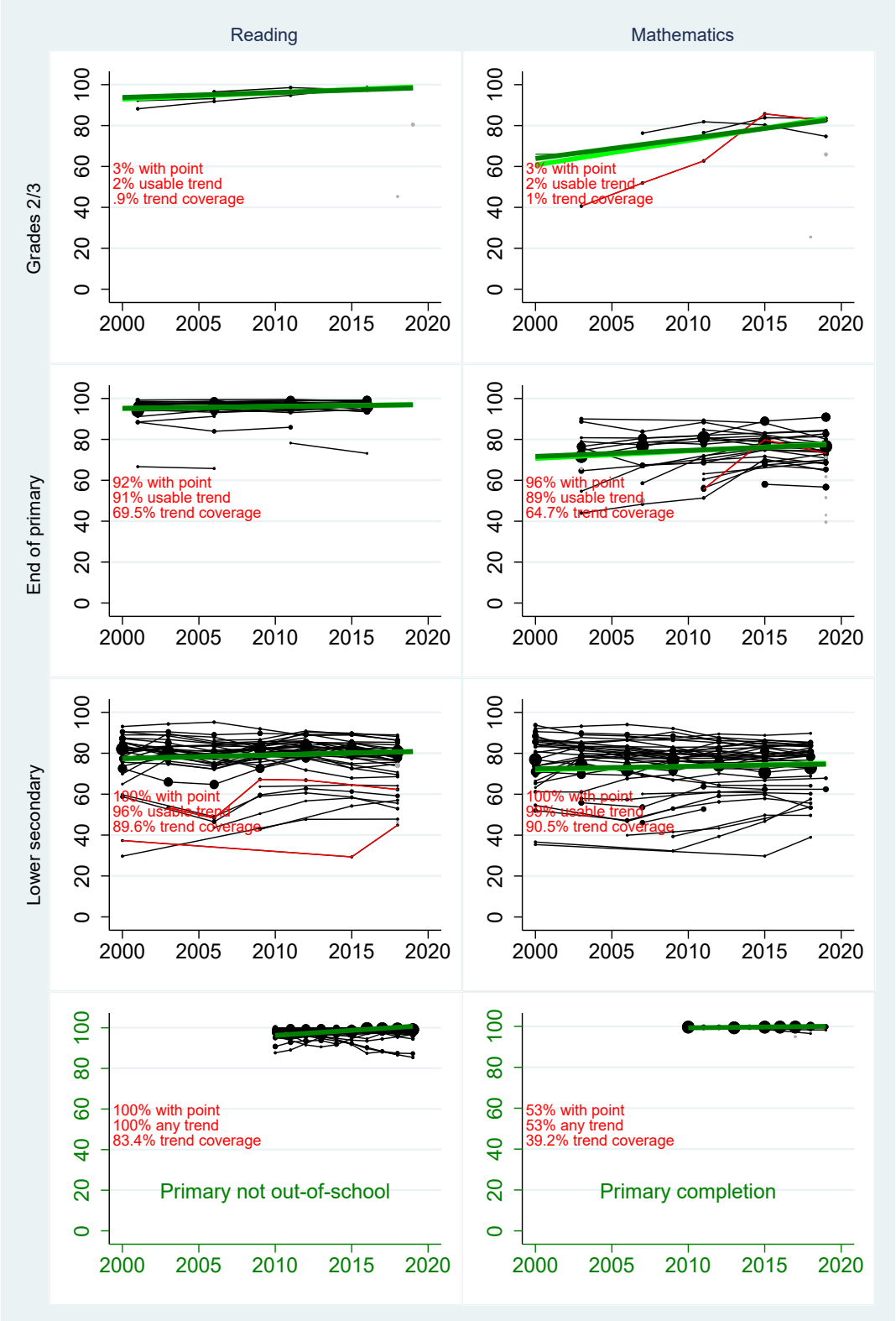
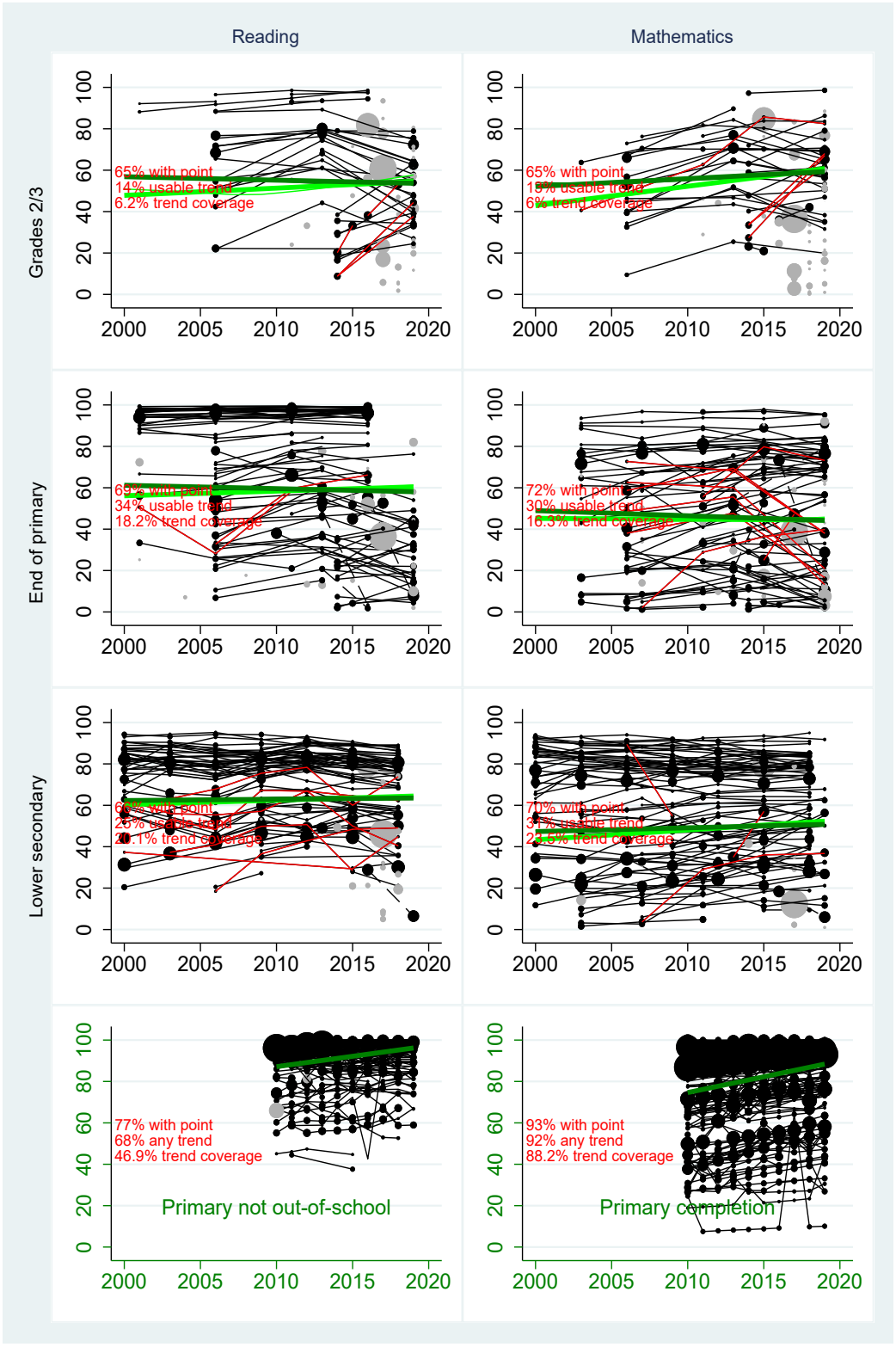
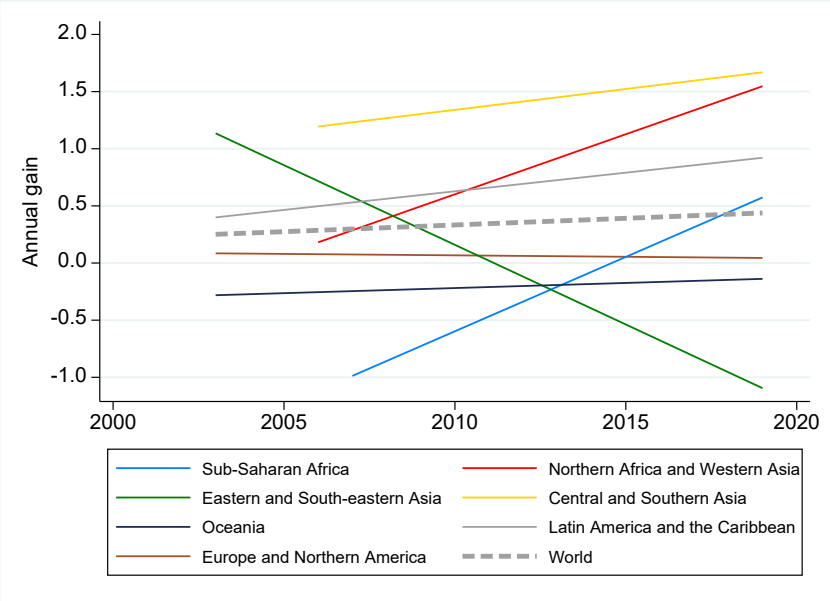


Figure 20: Global trends



The question can be asked whether there is a way of combining different levels and subjects when considering gains, partly to lessen the severity of the poor coverage of the SDG 4.1.1 statistics within each dimension. This may be feasible but is beyond the scope of the current report. However, Figure 21 provides some idea of what could be expected. This graph is like Figure 6 above, except here values have been weighted by country populations. The very steep decline for Eastern and South Eastern Asia is largely driven by a sudden, and rather unbelievable, decline in Indonesia's PISA reading scores between 2015 and 2018, at the lower secondary level.

Figure 13: Crude gain trends with no disaggregation and with LAC adjustments



11. Tables with regional and global trends

Table 8 provides statistics per world region. Here there is no separation between the six SDG 4.1.1 indicators, in other words by level and subject.

The column '% of children covered by the largest quarter of countries' draws from the kind of analysis presented in Section 7. According to the figures in this column, Oceania is the region with the greatest concentration of children in a few countries. The largest one quarter of countries, in terms of child populations, account for 95% of all children in the region. More specifically, Australia on its own accounts for 48% of the region's children. Even in the region with the lowest country-specific concentrations, Northern Africa and Western Asia, the largest quarter of countries account for 71% of all children.

The range in years for which 4.1.1 proficiency statistics are available in UIS.Stat is similar across regions.

Oceania is the region with the lowest percentage of children in countries with at least one proficiency statistic: the statistic for the region is 62%. This is largely because Papua New Guinea, which accounts for 31% of the region's children, has no 2000 to 2019 proficiency statistics in UIS.Stat. However, with respect to the availability of any trend data, regardless of consistency problems, Central and Southern Asia is the region that fares worst – here only 13% of children are in countries with some trend. The final column of the table indicates that virtually all the world's children are in countries where participation in schooling is measured.

The next table, Table 9, presents statistics per region, per 4.1.1 indicator, and with respect to the two participation indicators discussed in Section 10. The three columns whose values are represented within the graphs of Section 10 are the following: % of population with any data point; % of population with a usable trend; Trend coverage (%). The column '% of population with a usable trend' is the result of removing trends with inconsistent data sources across two years (see the discussion in Section 9).

The year range in Table 9 is the range of all statistics for the relevant SDG Indicator 4.1.1 whether they produce a trend or not.

The calculation of the 'Trend coverage' values was described in Section 10. These values provide a sense of how extensive the reporting on trends is within the year range of the first two columns of Table 9. The 'Average level within year range' is the average vertical height of the dark green trendlines in the Section 10 graphs. This means the restricted approach is used.

The two 'Annual percentage point gain' columns at the end of Table 9 present the slopes of the thick light green and dark green aggregate curves respectively of Section 10. The final annual gain statistic is the more conservative one, as it assumes that outside the range of years for which a country has statistics, the trend is flat.

It should be emphasized that gain statistics must be read while taking data coverage into account. Gain statistics where '% of population with a usable trend' is below 50% are especially likely to be biased by which countries have been excluded. The higher the coverage, the more the gain statistics of the last two column are likely to be reliable.

It is noteworthy that of the 42 rows of Table 9 representing the seven regions and six SDG 4.1.1 statistics, just three have a percentage of the population with a reliable trend exceeding 50% *and* have a trend that is greater than positive 0.50, using the continuous approach of the second-last column. An annual gain that is over positive 0.50 translates to one percentage point gain every two years. Two of these three relatively reliable and steep positive trends are in the region Latin America and the Caribbean. To illustrate, in this region 88% of children are in countries with Grades 2 to 3 mathematics trends, and the aggregate slope is positive 0.69, using the continuous approach. The upward trend can be seen in **Figure 17**. Compared to the three relatively reliable and steep *positive* trends, there are three relatively reliable and steep *negative* trends, using similar criteria. Two of these are in Latin America and the Caribbean (both at the end of primary), and one in Oceania. With regard to Latin America and the Caribbean, it is unlikely that so many countries in the region would experience the up and then down patterns demonstrated by the available SDG statistics. The contradictory patterns in this region are in fact largely a result of how LLECE statistics have been interpreted and adjusted (see **Appendix 4)**.

Given how important it is to establish the foundations for learning through reading in the early grades, what stands out is the low level of monitoring at this level, and for reading. It is among the least monitored of the six 4.1.1 indicators. Three regions – Northern Africa and Western Asia; Eastern and South-eastern Asia; and Central and Southern Asia – display no trend data for reading in the early grades.

In Table 9 some of the annual gain statistics are close to, or higher than 1.0, using the continuous approach, though this is never found in combination with a population with a usable trend exceeding 50%. Yet such gains are noteworthy, especially when they occur both in reading and mathematics. This is true for the lower primary gains in sub-Saharan Africa and end of primary gains in Central and Southern Asia – in both these regions and in both subjects annual 'continuous' gains of around 2.0 are seen. These statistics are based on around 30% of the child population in sub-Saharan Africa and around 12% in Central and Southern Asia[43].

The annual percentage point gains for reading at the end of primary in the sub-Saharan Africa region are negative in Table 9, despite substantial gains in this area being reported elsewhere[44]. The Table 9 patterns for this region, subject and level are largely explained by sharp downward trends in two large countries, Democratic Republic of the Congo (2010 to 2019) and United Republic of Tanzania (2006 to 2013). If these two countries are excluded, the slope for end of primary reading in sub-Saharan Africa becomes 0.97, instead of the negative 0.15 seen in Table 9 for the continuous approach.

As explained in Appendix 4, the 0.23 annual gain, using the continuous approach, for the world's reading at the end of primary becomes 0.39 when important adjustments are made (using the restricted approach, the outcome is an annual *loss* of 0.11 a year, slightly better than the 0.16 annual loss seen in Table 9).

[43] To illustrate, the mathematics trend is based on three countries in the Central and Southern Asia region: Bangladesh; Iran (Islamic Republic of); and Kazakhstan.
[44] UIS, 2021a.

Table 8: Overall breakdown by region

	Countries	Population aged 0 to 14 (thous.)	% of children covered by the largest quarter of countries	Year range for 4.1.1 statistics		% of population with any 4.1.1 data point	% of population with any 4.1.1 trend	% of population with any participation trend
Sub-Saharan Africa	52	478,492	73	2003	2019	93	67	100
Northern Africa and Western Asia	24	162,229	71	2000	2019	78	69	98
Central and Southern Asia	14	553,764	90	2001	2019	98	13	100
Eastern and South-eastern Asia	18	450,430	88	2000	2019	96	27	95
Oceania	21	10,261	95	2000	2019	62	58	98
Latin America and the Caribbean	48	155,744	90	2000	2019	92	88	100
Europe and Northern America	51	185,918	88	2000	2019	100	99	100
World	228	1,996,838	89	2000	2019	94	48	99

Table 9: Region breakdown with details by statistic

		Year range	% of population with any data point	% of population with any trend	% of population with a usable trend	Trend coverage (%)	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
Sub-Saharan Africa									
Grades 2/3 reading	2006	2019	75	29	29	5.7	32	1.94	0.43
Grades 2/3 mathematics	2011	2019	75	25	25	4.1	48	1.75	0.39
End of primary reading	2004	2019	69	60	44	16.5	32	-0.15	-0.18
End of primary mathematics	2006	2019	69	52	31	10.3	13	-0.47	-0.24
Lower secondary reading	2009	2019	19	10	0	0.0	-	-	-
Lower secondary mathematics	2003	2019	21	12	2	1.1	4	0.41	0.22
Primary aged not out-of-school	2010	2019	87	63		45.4	79	2.0	
Primary completion	2010	2019	100	99		95.6	59	2.7	
Northern Africa and Western Asia									
Grades 2/3 reading	2018	2019	3	0	0	0.0	-	-	-
Grades 2/3 mathematics	2018	2019	3	0	0	0.0	-	-	-
End of primary reading	2001	2016	30	17	17	9.4	54	0.14	0.13
End of primary mathematics	2003	2019	47	37	37	19.3	29	0.98	0.40
Lower secondary reading	2000	2018	42	22	22	15.1	60	0.49	0.33
Lower secondary mathematics	2000	2019	71	62	62	45.7	27	0.58	0.46

		Year range	% of population with any data point	% of population with any trend	% of population with a usable trend	Trend coverage (%)	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
Primary aged not out-of-school	2010	2019	88	88		77.3	91	1.0	
Primary completion	2010	2019	90	88		77.3	86	1.1	
Central and Southern Asia									
Grades 2/3 reading	2016	2019	92	0	0	0.0	-	-	-
Grades 2/3 mathematics	2016	2019	92	0	0	0.0	-	-	-
End of primary reading	2001	2017	96	12	12	4.3	51	2.42	0.36
End of primary mathematics	2003	2019	96	13	13	4.0	37	2.39	0.70
Lower secondary reading	2009	2018	75	9	9	1.7	50	2.24	0.27
Lower secondary mathematics	2003	2019	78	13	13	3.8	42	3.32	0.94
Primary aged not out-of-school	2010	2019	82	75		30.9	97	0.7	
Primary completion	2010	2019	99	99		99.1	83	1.3	
Eastern and South-eastern Asia									
Grades 2/3 reading	2011	2019	61	0	0	0.0	-	-	-
Grades 2/3 mathematics	2011	2019	61	0	0	0.0	-	-	-
End of primary reading	2001	2019	35	16	16	5.0	62	2.37	0.85
End of primary mathematics	2003	2019	39	2	2	0.9	96	-0.12	-0.03
Lower secondary reading	2000	2018	91	27	27	20.8	57	-0.40	-0.22
Lower secondary mathematics	2000	2019	92	27	27	20.9	48	-0.29	-0.11
Primary aged not out-of-school	2010	2019	37	37		27.8	96	0.8	
Primary completion	2010	2019	95	93		93.4	94	1.2	
Oceania									
Grades 2/3 reading	2011	2019	53	49	49	14.7	93	0.31	0.10
Grades 2/3 mathematics	2003	2019	53	49	49	41.6	69	0.28	0.26
End of primary reading	2001	2016	12	9	9	7.3	91	-0.01	-0.01
End of primary mathematics	2003	2019	61	58	58	49.4	63	0.15	0.14
Lower secondary reading	2000	2018	58	58	58	55.2	85	-0.39	-0.38
Lower secondary mathematics	2000	2018	58	58	58	55.2	83	-0.67	-0.67
Primary aged not out-of-school	2010	2019	98	67		60.8	97	0.6	
Primary completion	2010	2019	85	35		35.0	82	1.7	
Latin America and the Caribbean									
Grades 2/3 reading	2006	2019	92	88	88	60.1	70	-0.25	-0.17
Grades 2/3 mathematics	2006	2019	92	88	88	60.1	60	0.69	0.51

		Year range	% of population with any data point	% of population with any trend	% of population with a usable trend	Trend coverage (%)	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
End of primary reading	2001	2019	91	88	88	60.2	46	-0.51	-0.38
End of primary mathematics	2006	2019	91	88	88	60.1	43	-1.10	-0.82
Lower secondary reading	2000	2018	86	75	75	65.6	50	0.29	0.28
Lower secondary mathematics	2000	2019	87	75	75	65.4	32	0.54	0.49
Primary aged not out-of-school	2010	2019	97	95		87.8	97	0.6	
Primary completion	2010	2019	100	99		84.9	93	0.8	
Europe and Northern America									
Grades 2/3 reading	2001	2019	3	2	2	0.9	96	0.31	0.24
Grades 2/3 mathematics	2003	2019	3	2	2	1.0	73	1.20	0.98
End of primary reading	2001	2016	92	91	91	69.5	96	0.10	0.09
End of primary mathematics	2003	2019	96	89	89	64.7	75	0.37	0.29
Lower secondary reading	2000	2018	100	96	96	89.6	79	0.18	0.18
Lower secondary mathematics	2000	2019	100	99	99	90.5	74	0.17	0.11
Primary aged not out-of-school	2010	2019	100	100		83.4	99	0.5	
Primary completion	2010	2019	53	53		39.2	100	0.1	
World									
Grades 2/3 reading	2001	2019	65	14	14	6.2	55	0.35	-0.17
Grades 2/3 mathematics	2003	2019	65	13	13	6.0	56	0.93	0.38
End of primary reading	2001	2019	69	38	34	18.2	60	0.23	-0.16
End of primary mathematics	2003	2019	72	35	30	16.3	47	-0.04	-0.25
Lower secondary reading	2000	2019	66	28	25	20.1	63	0.22	0.06
Lower secondary mathematics	2000	2019	70	33	31	23.5	49	0.47	0.18
Primary aged not out-of-school	2010	2019	77	68		46.9	92	1.0	
Primary completion	2010	2019	93	92		88.2	82	1.5	

The proficiency level statistics provided in the third-last column of Table 9 are based on countries with statistics for a specific indicator from more than one point in time. Level and annual gain statistics are based on the same set of countries. Given that only 13% to 34% of the world's children, depending on which of the six proficiency indicators is being considered, are in countries with the required trend statistics, the question arises how this biases the *level* statistics. **Table 10** reflects this bias, specifically how much higher the proficiency level would be if countries *without* trends were included. For example, the level of reading proficiency at the end of primary in sub-Saharan Africa is 32% in Table 9 but one percentage point higher, or 33%, if even countries with no trends are included (and the statistic is weighted by each country's child population). Here the difference is small.

Table 10: Change to proficiency levels when all data are considered

	Grades 2/3		End of primary		Lower secondary	
	Read.	Math.	Read.	Math.	Read.	Math.
Sub-Saharan Africa	-3	-13	1	9		16
Northern Africa and Western Asia			8	-1	-15	0
Central and Southern Asia			-10	-4	-4	-25
Eastern and South-eastern Asia			-16	-62	10	15
Oceania	-2	-1	-8	2	0	0
Latin America and the Caribbean	-1	1	-2	-3	-2	-2
Europe and Northern America	-9	-4	0	-1	0	0
World	0	-9	-12	-11	-7	-8

However, there are some large differences. Above all, in the region Eastern and South-eastern Asia, proficiency in end of primary mathematics is 96% in Table 9 but only 34% if countries with no trend statistics are included. The 96% is based on just three locations – Republic of Korea, Hong Kong and Singapore – while the 34% value emerges when several other large countries with statistics from just one point in time, including Indonesia and Philippines, are brought in. Similarly, in Central and Southern Asia the lower secondary mathematics value in Table 9 drops if India is brought in, while in sub-Saharan Africa the lower secondary mathematics value in Table 9 rises if Uganda, Rwanda and a few other countries are brought in.

At the global level, it is clear that in general Table 9 biases levels upwards – none of the values in the final row of Table 10 are positive. Countries with less data tend to have lower proficiency levels, as already discussed in Section 6.

12. Disaggregation by gender

Table 11 provides a breakdown using sex for key statistics from Table 8, while **Table 12** does the same for Table 9. According to Table 11, statistics by sex are less available than statistics for both sexes combined in two regions: sub-Saharan Africa and Eastern and South-eastern Asia. In addition, Table 12 reveals the region Central and Southern Asia has limited data by sex: here for end of primary the data coverage is low relative to what was seen in Table 9.

Table 11: Overall breakdown by region and sex

	% of population with any 4.1.1 trend	
	Both sexes combined	By sex
Sub-Saharan Africa	67	45
Northern Africa and Western Asia	69	69
Central and Southern Asia	13	13
Eastern and South-eastern Asia	27	20
Oceania	58	58
Latin America and the Caribbean	88	88
Europe and Northern America	99	99
World	48	41

A remarkable pattern seen in Table 12 is that among the 17 regional instances where males and females do *not* display the same proficiency across both subjects at a specific school level, it is nearly always the case that reading allows females to gain an advantage, while mathematics allows males to gain an advantage. There is just one exception to this pattern: in Eastern and South-eastern Asia the female advantage is driven by high performance not just in reading, which is common, but also mathematics. At the global level, females perform better in reading than males at all three school levels, and males perform better in mathematics than females.

Moreover, females outperform males insofar as there are 18 region-specific rows in Table 12 where the female level of proficiency is higher than that for males, against 12 rows where the opposite applies.

There is only one region where the *participation* of males unequivocally surpasses that of females: Northern Africa and Western Asia. In this region females lag very slightly behind males with respect to both the out-of-school and completion indicators. This could explain the proficiency advantage for females in this region. This would be the case if a substantial number of poorly performing females were excluded from schooling. However, if the rest of the world is considered, where there is no clear participation advantage for males, the statistics do *not* suggest that participation is the factor explaining better female proficiency. Instead, the statistics would be compatible with the hypothesis that females learn better at school.

Turning to trends, there are up to twelve annual gain statistics per region and per sex. In three regions these statistics mostly favour females: sub-Saharan Africa; Northern Africa and Western Asia; and Eastern and South-eastern Asia. In these three regions females appear to be advancing faster than males, despite often have a higher point of departure. In the remaining regions, except for Oceania, more trend statistics favour males. One pattern that stands out is an exceptionally large gap in favour of male progress in Central and Southern Asia with respect to both subjects at the lower secondary level. Using the continuous approach, males were advancing faster than females by a whole percentage point in reading. Behind this trend are just two countries, both of which display relatively fast progress for males: Bangladesh and Kazakhstan. In Oceania trend statistics are fairly balanced across the two sexes. Globally, the gain statistics are not strongly favouring males or females, though males appear to be progressing faster at the lower primary level, while the trend statistics favour females at the end of primary level.

Table 12: Region breakdown with details by sex-specific statistic

	% of population with a usable trend by sex	Average level within year range		Annual percentage point gain (continuous approach)		Annual percentage point gain (restricted approach)	
		Males	Females	Males	Females	Males	Females
Sub-Saharan Africa							
Grades 2/3 reading	25	33	34	2.32	2.35	0.45	0.48
Grades 2/3 mathematics	25	50	45	1.71	1.73	0.37	0.42
End of primary reading	32	33	36	0.44	0.71	0.08	0.19
End of primary mathematics	28	15	12	-0.60	-0.24	-0.34	-0.14
Lower secondary reading	0	-	-	-	-	-	-
Lower secondary mathematics	2	5	3	0.48	0.33	0.26	0.18
Primary aged not out-of-school	60	81	76	2.1	2.1		
Primary completion	99	59	60	2.8	2.7		
Northern Africa and Western Asia							
Grades 2/3 reading	0	-	-	-	-	-	-
Grades 2/3 mathematics	0	-	-	-	-	-	-
End of primary reading	17	46	60	-0.34	-0.25	-0.20	-0.12
End of primary mathematics	37	29	29	0.94	1.04	0.37	0.43
Lower secondary reading	22	52	68	0.52	0.38	0.38	0.20
Lower secondary mathematics	62	27	27	0.49	0.68	0.38	0.54
Primary aged not out-of-school	49	89	88	1.5	1.0		
Primary completion	88	87	85	1.2	1.3		
Central and Southern Asia							
Grades 2/3 reading	0	-	-	-	-	-	-
Grades 2/3 mathematics	0	-	-	-	-	-	-
End of primary reading	4	60	71	0.77	1.21	0.70	1.09
End of primary mathematics	5	35	35	1.36	1.11	1.26	1.05
Lower secondary reading	9	49	51	2.74	1.75	0.35	0.19
Lower secondary mathematics	13	45	38	3.51	3.09	0.98	0.89
Primary aged not out-of-school	72	96	98	1.1	0.4		
Primary completion	99	83	83	1.4	1.4		
Eastern and South-eastern Asia							
Grades 2/3 reading	0	-	-	-	-	-	-
Grades 2/3 mathematics	0	-	-	-	-	-	-
End of primary reading	16	57	67	2.42	2.26	0.87	0.81
End of primary mathematics	2	96	97	-0.15	-0.09	-0.04	-0.02
Lower secondary reading	20	42	56	-0.12	-0.10	-0.12	-0.10
Lower secondary mathematics	20	39	39	-0.18	0.11	-0.18	0.10
Primary aged not out-of-school	28	95	95	0.9	1.1		
Primary completion	93	93	95	1.1	1.1		
Oceania							
Grades 2/3 reading	49	92	95	0.38	0.27	0.12	0.08
Grades 2/3 mathematics	49	69	68	0.28	0.29	0.25	0.27

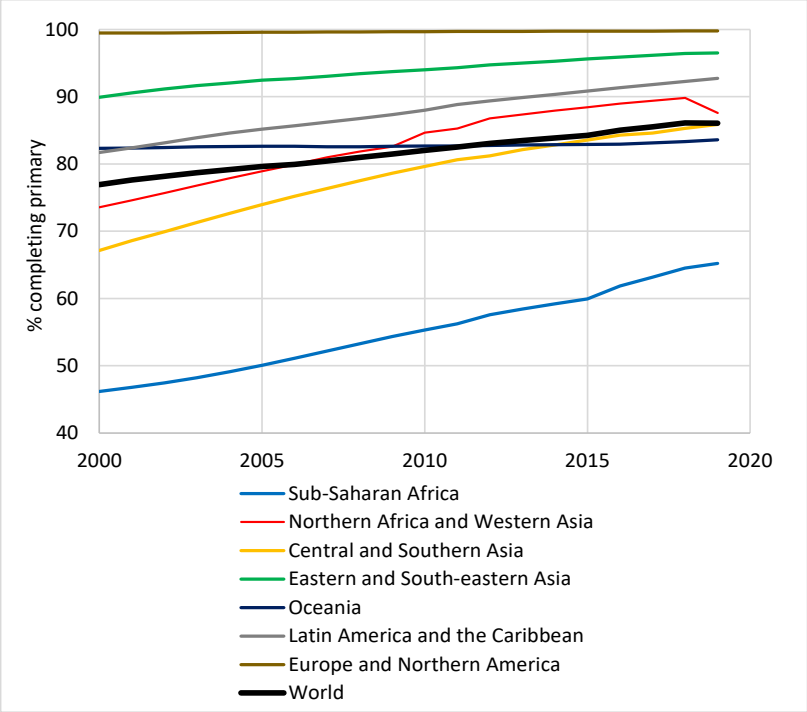
	% of population with a usable trend by sex	Average level within year range		Annual percentage point gain (continuous approach)		Annual percentage point gain (restricted approach)	
		Males	Females	Males	Females	Males	Females
End of primary reading	9	89	93	0.04	-0.06	0.03	-0.06
End of primary mathematics	58	65	62	-0.03	0.32	-0.03	0.29
Lower secondary reading	58	80	90	-0.43	-0.37	-0.42	-0.36
Lower secondary mathematics	58	83	82	-0.67	-0.69	-0.66	-0.68
Primary aged not out-of-school	65	97	96	0.4	0.3		
Primary completion	35	81	83	2.4	1.0		
Latin America and the Caribbean							
Grades 2/3 reading	88	67	73	-0.23	-0.35	-0.16	-0.25
Grades 2/3 mathematics	88	60	60	0.74	0.59	0.54	0.43
End of primary reading	88	44	49	-0.57	-0.46	-0.42	-0.34
End of primary mathematics	88	45	42	-1.17	-1.06	-0.87	-0.80
Lower secondary reading	75	44	55	0.31	0.28	0.32	0.28
Lower secondary mathematics	75	35	29	0.53	0.52	0.49	0.48
Primary aged not out-of-school	58	95	96	1.0	1.1		
Primary completion	99	92	94	0.9	0.8		
Europe and Northern America							
Grades 2/3 reading	2	95	97	0.40	0.21	0.30	0.17
Grades 2/3 mathematics	2	74	72	1.15	1.25	0.93	1.04
End of primary reading	83	95	97	0.16	0.05	0.13	0.04
End of primary mathematics	82	75	73	0.43	0.32	0.32	0.28
Lower secondary reading	96	74	84	0.28	0.11	0.28	0.11
Lower secondary mathematics	99	74	73	0.26	0.21	0.15	0.12
Primary aged not out-of-school	93	98	98	0.6	0.6		
Primary completion	53	100	100	0.1	0.1		
World							
Grades 2/3 reading	13	56	59	0.40	0.29	-0.16	-0.25
Grades 2/3 mathematics	13	57	55	0.97	0.84	0.41	0.32
End of primary reading	29	60	65	0.12	0.16	-0.10	-0.08
End of primary mathematics	26	48	46	-0.36	-0.31	-0.34	-0.30
Lower secondary reading	24	56	67	0.4	0.22	0.16	0.05
Lower secondary mathematics	29	49	46	0.57	0.57	0.19	0.23
Primary aged not out-of-school	58	91	90	1.3	1.2		
Primary completion	92	81	82	1.6	1.6		

13. Combining proficiency and participation statistics

In order to explore trends in the percentage of *children in the population* who are proficient, a newly produced dataset of school completion of UNESCO was used. This dataset, which is a product of imputations and corrections explained in Dharamshi *et al* (2021), has annual primary completion values over around thirty years for 164 countries, or 98.4% of the global child population – substantial missing countries are Saudi Arabia, Sri Lanka and Democratic People's Republic of Korea. The statistics for all countries begin in 1990, and for 90% of children they reach at least 2019 (for 97% they reach at least 2015). The dataset was retrieved from the VIEW[45] portal of UNESCO[46]. The primary completion rates in the data reflect completion in the population aged three to five years above the official end of primary age[47].

The historical patterns for 2000 to 2019 seen in this dataset are shown in **Figure 20**. Trends are weighted by the country-specific annual child population aged 10 to 14. The global increase during 2010 to 2019, from 82% to 86%, is less steep than what was seen in earlier Figure 19. Yet the changes are substantial, a key advantage with the VIEW statistics being that they extend before 2010. The Figure 20 improvement for the entire 2000 to 2019 period is from 77% to 86% globally.

Figure 14: Completion of primary schooling



[45] Visualizing Indicators of Education for the World.

[46] Retrieved November 2022, https://education-estimates.org. Just the primary completion rates (the 'CR Indicator' version) were retrieved.

[47] While the new dataset provides what seems to be the best global data permitting comparison across countries and time, it is not necessarily the most reliable source for individual countries. It relies to a large extent on international data collection initiatives, such as the Demographic and Health Survey, and only to a limited extent on national household survey programmes. Insofar as the latter provide more reliable level and trend statistics, the global dataset would be of limited use for national planners.

Table 13 is meant to be compared to the end of primary values of earlier Table 9. For Table 13, proficiency statistics used were just those where there was a VIEW completion statistic for the same country and year. This explains why, for instance, 32% of the world's children had usable data for end of primary reading in Table 13, against 34% in Table 9. Within each country with the required trend data, the level of proficiency *in the child population* was calculated. For instance, if 40% of enrolled children were proficient, and 80% of children completed primary schooling, then 32% of the population (40% times 80%) were assumed to be proficient. The assumption of UIS (2017b) was thus used that children who do not complete primary schooling do not display the learning proficiency expected of children who complete primary. A region such as sub-Saharan Africa, with relatively low levels of primary completion, even in more recent years, would display a large gap between the enrolment- and population-based levels of proficiency: for reading the two statistics would be 32% (Table 9) and 23% (Table 13).

The percentage point gain statistics shown in Table 13 refer to the percentage of an age cohort of the child population, not the percentage of enrolled children. Thus, for instance, a percentage point gain in sub-Saharan Africa in Table 13 would refer to more children than a percentage point gain in Table 9. Importantly, the 0.33 percentage point annual gain for reading and the world, using the 'continuous approach', becomes 0.53 when the adjustments described in Appendix 4 are applied. This is reflected in Figure 24. What does this 0.53 mean? Roughly, it means that the percentage of proficient children increased by around eleven percentage points over the twenty-year period. The 0.53 is half the 'almost a percentage point per year' put forward speculatively at the end of section 10. This can be entirely explained by the less steep completion improvements seen in Figure 20 compared to Figure 19. It can be assumed that the slower improvements in completion seen in Figure 20 are more reliable as they are the result of a detailed exercise to impute and correct statistics. The decline at the right-hand end of the Northern Africa and Western Asia trend seen in Figure 20 is due to a major country with high primary completion, Turkey, not having a 2019 value. Completion statistics behind Figure 20 have values for Syrian Arab Republic only up to 2011, meaning school participation disruptions due to the war in that country would not be reflected. In the case of Syrian Arab Republic, there are no SDG proficiency trends beyond 2011, meaning the absence of completion statistics would have no effect on Table 13. However, the absence of 2019 completion statistics for Turkey would flatten the region's end of primary and lower secondary trends a little as Turkey's improvements in Grades 4 and 8 between 2015 and 2019 in TIMSS would not be used for Table 13 (though these improvements would influence the statistics in Table 9).

Of note is the fact that an apparent deterioration in end of primary reading in sub-Saharan Africa seen in Table 9 – see the 0.15 annual percentage point *decline* – becomes an improvement when participation is taken into account, as in Table 13 with its 0.14 annual *gain*.

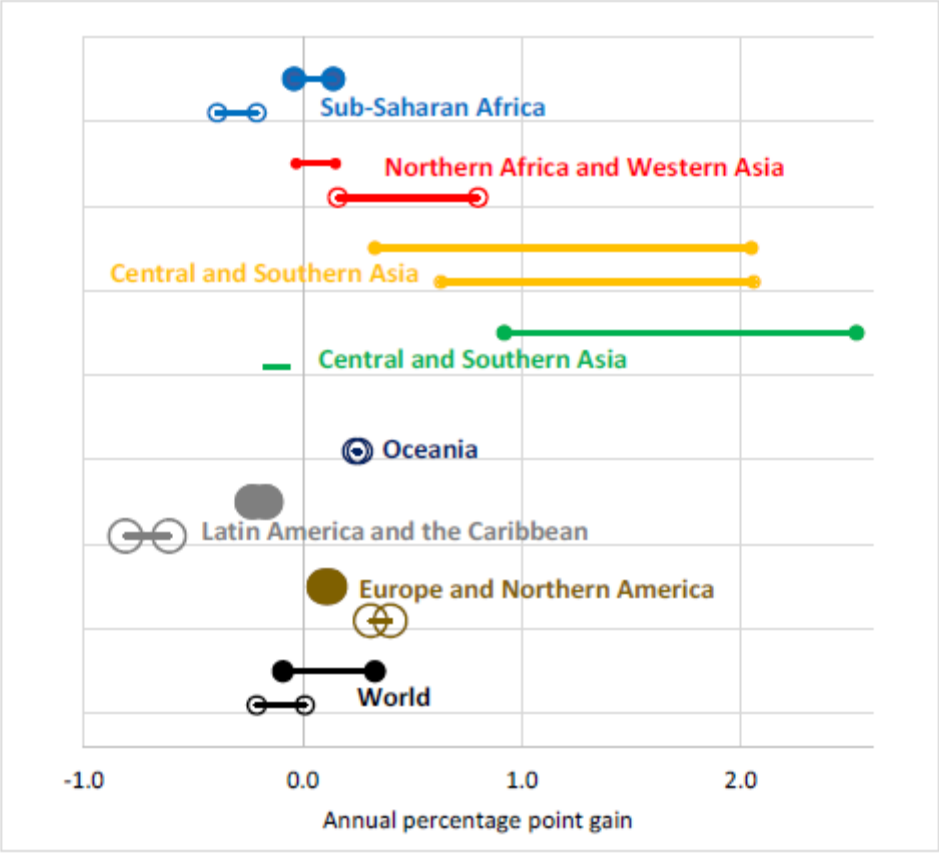
Table 13: Population-focused region breakdown with end of primary details

	Year range		% of population with a usable trend	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
Sub-Saharan Africa						
End of primary reading	2004	2019	39	23	0.14	-0.04
End of primary mathematics	2006	2019	26	8	-0.39	-0.21
Northern Africa and Western Asia						
End of primary reading	2001	2016	8	32	-0.03	0.15
End of primary mathematics	2003	2019	29	27	0.80	0.16
Central and Southern Asia						
End of primary reading	2001	2017	12	43	2.05	0.33
End of primary mathematics	2003	2019	13	31	2.06	0.63
Eastern and South-eastern Asia						
End of primary reading	2006	2019	16	57	2.53	0.92
End of primary mathematics	2007	2019	1	97	-0.17	-0.07
Oceania						
End of primary reading	-	-	-	-	-	-
End of primary mathematics	2003	2019	49	63	0.26	0.24
Latin America and the Caribbean						
End of primary reading	2001	2019	87	42	-0.23	-0.17
End of primary mathematics	2006	2019	87	40	-0.81	-0.61
Europe and Northern America						
End of primary reading	2001	2016	91	96	0.12	0.10
End of primary mathematics	2003	2019	86	74	0.40	0.31
World						
End of primary reading	2001	2019	32	55	0.33	-0.09
End of primary mathematics	2003	2019	28	45	0.01	-0.21

Figure 23 draws largely from the last two columns of Table 13. The predominance of slopes greater than zero at the region level is clear. Only one region has both continuous slope statistics below zero: Latin America and the Caribbean. However, the reliability of these trends in this region are questionable, as discussed in Section 11 (and **Appendix 4**).

Figure 15: Population-focused region breakdown with end of primary details

Figure 21: Population-focussed region breakdown with end of primary details



Note: Solid markers reflect reading, markers with no fill reflect mathematics. The lower (left-hand) marker uses the 'restricted approach' while the higher (right-hand) marker uses the 'continuous approach'. Marker areas are proportional to '% of population with a usable trend'.

If three adjustments discussed in Appendix 4 relating to the Latin America and the Caribbean region are applied to the UIS.Stat input data, the result is Figure 24 below. Here all four slopes for this region become positive, and the global statistics are shifted to the right.

Figure 16: Previous graph with LAC adjustments

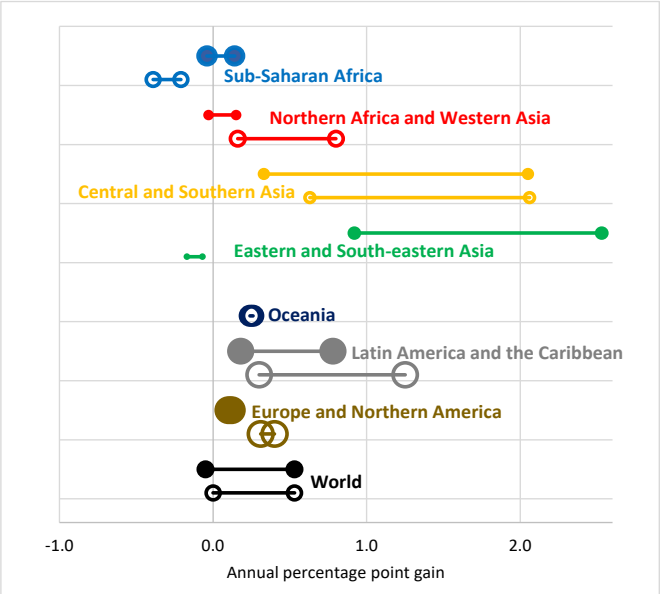


Figure 25 is the result of using the following slopes and average levels: the slopes are those from the continuous approach seen in Figure 24 above; the levels are those that *would* emerge in the third-last column of Table 9, *if* the three adjustments for Latin America and the Caribbean described in Appendix 4 are applied. Figure 25 focusses on end of primary reading. This is perhaps the most reliable picture that can be obtained of proficiency in the *population* – end of primary reading is relatively well covered in the data, it is a level that is freer from measurement complexities than lower primary grades, and here important adjustments for one region have been implemented.

Figure 17: Population-focussed trends for end of primary reading

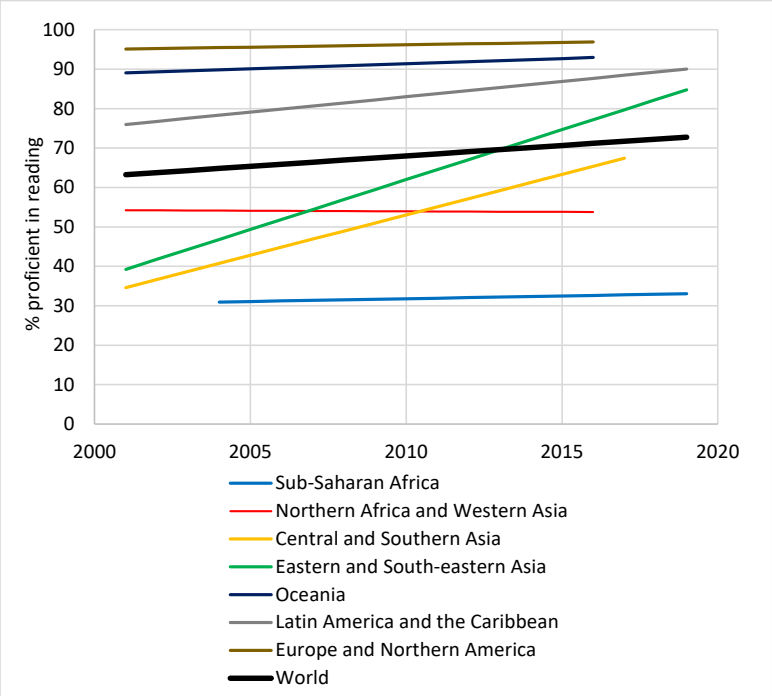
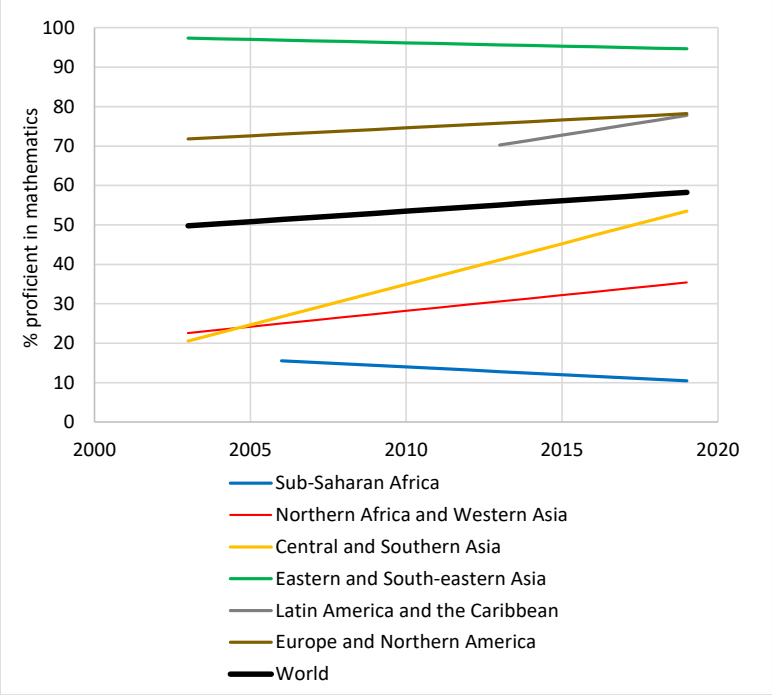


Figure 26 below is like the previous Figure 25, except here mathematics is the focus.

Figure 18: Population-focussed trends for end of primary mathematics



14. The way forward for regional and global monitoring of proficiency

The title of the report includes the following question: How close are we to reliable regional and global SDG 4.1.1 trend statistics? The analysis of the preceding sections is in part aimed at answering this question.

If the standard is that all the world's children should be in countries where proficiency trends have been measured at some point since 2000, and these trends are reported on in the global UNESCO reporting system, then the world is around halfway to the ideal. Some 52% of children are not in such countries. What greatly facilitates progress towards the measurement ideal is that 31% of the world's children live in India or China. The availability of measures from these two countries would reduce the 52% unmeasured to around 21%, who would be children spread across some 104 countries. Addressing this gap of one-fifth should be considered a substantial challenge, given how many countries are involved.

It could be argued that the above standard is unambitious as it ignores the fact that in the period since 2000 several trend measurements are rather old. While this is a valid concern, it is clear that a rising percentage of the world's children have become subject to the measurement of proficiency trends. The figure has risen more or less continuously from 20% in 2000 to around 40% at the start of the pandemic[48].

[48] Figure 2.

However, the focus should not just be on the availability of statistics, but also on improvements in their quality. Here it is difficult to draw hard conclusions. There appears to be an increasing volume of sufficiently reliable statistics over time, but whether the average quality is improving would be difficult to tell, even with further analysis. This report has pursued a few approaches to validating the quality of existing data, such as examining the internal coherence of test microdata (Appendix 3) and evaluating the completeness of existing metadata (Section 5).

Progress with respect to data coverage and quality can be said to rest on two preconditions *at the country level*. National government and non-governmental organizations have a vital role to play in facilitating a robust global monitoring system. First, there needs to be sufficient awareness of the importance of foundational skills of the kind measured by SDG 4.1.1 for national development and resilience of society to multiple economic, technological and environment challenges. This should include an awareness that progress with respect to these skills has tended to be slow, meaning the difference between success and no success in the education sector can be small, and requires very careful measurement. Secondly, within-country capacity to design and implement systems that monitor the proficiency levels of children must be vastly improved. In the absence of these preconditions, there will not be sufficient interest in a comprehensive global monitoring system, and the critical engagement that is necessary to raise the reliability of statistics will be lacking.

Global organizations, including the UIS, can do more to advance these two preconditions, by promoting a deeper and society-wide understanding of the importance of foundational skills and building capacity through methodological innovation, the documentation of interesting country case studies, and training in the relevant areas of statistics. A lot has been done. For instance, the Rosetta Stone initiative has facilitated better comparison across international testing programmes at any point in time[49]. However, as argued in UIS (2018), innovation should increasingly shift to the measurement of trends over time, which in many respects is what countries are most interested in. The pandemic prompted some innovation in this regard, with the running of national and international tests in parallel, partly with a view to verifying the ability of national assessments to detect change[50].

Country-specific case studies are important in part because they can highlight certain tensions between, regional and global monitoring and national monitoring. These tensions are often not sufficiently understood. The global datasets on SDG monitoring are, at least in the case of learning proficiency, aimed at producing, within certain resource constraints, global datasets where across-country and across-time comparisons are as credible as possible. This does not mean the statistics for individual countries are the most reliable statistics available, or possible. There is simply not the capacity within, for instance, UNESCO to examine and properly evaluate all the data of relevance in every one of over two hundred countries. Thus, UNESCO's systems must inevitably rely to a large extent on what is readily available in large international testing systems, and in the case of participation, harmonized datasets with considerable use of imputation (see Section 13). Good case studies could assist in the understanding of how global and country-level monitoring could be aligned better, while taking into account cost constraints. This in turn could assist national analysts to understand better the meaning of their national statistics within UIS.Stat, but also to focus on improving national statistical systems.

[49] UIS, 2022b; UIS, 2022c.
[50] UIS, 2022a.

Reporting on the available proficiency statistics by the UIS has improved, apart from the wider coverage of the data. Metadata, specifically the source of every published value, is now more easily seen in the online system. Two further improvements have been a topic of discussion, and they seem to warrant serious attention. One is more details on the degree and type of country involvement with respect to every proficiency statistic. While statistics from international testing programmes imply some country-level involvement in the programme, there are cases where countries have requested that the statistics not be used for SDG reporting. Some details on this should be available. Moreover, a country may prefer a statistic from one international programme to be used over a statistic from some other such programme. And assuming national programmes increasingly come to be seen as a legitimate source for reporting, why a country would want this to be used should be made clear.

The second improvement relates to the integration of within-school proficiency statistics and out-of-school statistics. This integration could be implemented programmatically in the online reporting system, along the lines of what has been presented in this report, yet difficult decisions would need to be taken. Should official SDG statistics on participation (such as those used for Sections 10 and 11) be used, or statistics from a more internally consistent global dataset such as VIEW (see Section 13)? How should country inputs be sought in such a process, bearing in mind that reliability problems in the two separate measures could compound each other within the combined indicator? As a minimum, and this would not be difficult, warnings that typical SDG Indicator 4.1.1 values do not take into account changes in participation should be made clearer. As shown in this report, taking participation into account can make a substantive difference to level and trend statistics.

References

ASER Centre (2019). *ASER - Learning Trends (2012-2018): All India.* New Delhi.

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

Azevedo, J.P., Akmal, M., Cloutier, M., Rogers, H. & Wong, Y.N. (2022). *Learning losses during COVID-19: Global estimates of an invisible and unequal crisis.* Washington: World Bank.

Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W.H. & Zakharov, A. (2016). Revisiting the relationship between international assessment outcomes and educational production: Evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal*, 53(4): 1054-85.

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.

Dang, H., Glewwe, P., Lee, J., and Vu, K. (2020). What explains Vietnam's exceptional performance in education relative to other countries? Analysis of the 2012 and 2015 PISA data. Oxford: RISE.

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

Flotts, M.P., Manzi, J, Jiménez, D. & Abarzúa, A. (2015). *Informe de resultados TERCE: Logros de aprendizaje*. Santiago: UNESCO.

Gustafsson, M. (2020). A revised PIRLS 2011 to 2016 trend for South Africa and the importance of analysing the underlying microdata. Stellenbosch: Stellenbosch University.

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.

Johnson, D. & Parrado, A. (2021). Assessing the assessments: Taking stock of learning outcomes data in India. *International Journal of Educational Development*, 84.

Martin, M.O., von Davier, M. & Mullis, I.V.S., eds. (2020). *Methods and procedures in TIMSS 2019*. Chestnut Hill: IEA.

Moscoviz, L. & Evans, D.K. (2022). *Learning loss and student dropouts during the COVID-19 pandemic: A review of the evidence two years after schools shut down*. Washington: Center for Global Development.

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

Pacific Community (2019). *Pacific Islands literacy and numeracy assessment 2018: Regional report*. Suva.

Ross, K.N., Saito, M., Dolata, S. & Ikeda, M. (2008). *Chapter 2: The conduct of the SACMEQ II project*. Paris: IIEP.

UIS (2017a). *Principles of Good Practice in Learning Assessment*. Montreal.

UIS (2017b). *Counting the number of children not learning: Methodology for a global composite indicator for education*. Montreal.

UIS (2018). *Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1)*. Montreal.

UIS (2019). *How fast can levels of proficiency improve? Examining historical trends to inform SDG 4.1.1 scenarios*. Montreal.

UIS (2020). *Evidence-based projections and benchmarks for SDG Indicator 4.1.1*. Montreal.

UIS (2021a). *Assessing learning proficiency levels and trends for Sustainable Development Goal 4.1: A focus on Africa.* Montreal.

UIS (2021b). *Pandemic-related disruptions to schooling and impacts on learning proficiency indicators: A focus on the early grades*. Montreal.

UIS (2022a). *COVID-19 in Sub-Saharan Africa: Monitoring impacts on learning outcomes.* Montreal.
UIS (2022b). *Rosetta Stone analysis report: Establishing a concordance between ERCE and TIMSS/PIRLS.* Montreal.
UIS (2022c). *Analysis report: Establishing a concordance between PASEC and TIMSS/PIRLS.* Montreal.
UNESCO (2005). *Education for All global monitoring report 2005: The quality imperative*. Paris.
UNESCO (2014). *Primera entrega de resultados TERCE*. Santiago.
UNESCO (2016). *Reporte técnico: Tercer Estudio Regional Comparativo y Explicativo*. Santiago.
UNESCO (2017). *Revised medium-term strategy 2017-2021 of the UNESCO Institute for Statistics (UIS)*. Montreal.
UNESCO (2021). *Los aprendizajes fundamentales en América Latina y el Caribe: Evaluación de logros de los estudiantes: Resumen ejecutivo.* Paris.
World Bank (2019). *Ending learning poverty: What will it take?* Washington.
World Bank (2022). *The state of global learning poverty: 2022 update*. Washington.

Appendix 1: Trends by World Bank country income category

The following three tables provide breakdowns by World Bank country income categories, as calculated by the World Bank for 2022. The methods are those explained in earlier Sections 11 and 13.

The warnings about annual gain (and proficiency level) statistics expressed in Section 11 in relation to Table 9 apply here too. To illustrate, the closeness of the level statistics for Grades 2 and 3 mathematics across the two categories low-income countries and lower middle income countries in Table 15 is very unlikely to be real. The fact that only 7% of the child population in the latter category has trend data makes the 50% proficiency level here particularly unlikely to be a true reflection of reality.

Importantly, if the region Latin America and the Caribbean is excluded when Table 15 is produced, the four negative gain statistics for the category upper middle-income countries all become clearly positive, and are all at least positive 0.5. However, the reliability of these more favourable statistics is low, with '% of population with a usable trend' being around 10%. These patterns reflect the problem with the Latin America and the Caribbean trends discussed in Section 11.

Table 14: Overall breakdown by World Bank income group

	Countries	Population aged 0 to 14 (thous.)	% of children covered by the largest quarter of countries	Year range for 4.1.1 statistics		% of population with any 4.1.1 data point	% of population with any 4.1.1 trend	% of population with any participation trend
Low income countries	28	296,867	59	2003	2019	89	72	98
Lower middle income countries	54	986,780	86	2000	2019	96	34	100
Upper middle income countries	53	505,535	92	2000	2019	96	45	99
High income countries	77	199,311	88	2000	2019	92	91	92

Table 15: World Bank income group breakdown with details by statistic

	Year range		% of population with any data point	% of population with any trend	% of population with a usable trend	Trend coverage (%)	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
Low income countries									
Grades 2/3 reading	2012	2019	54	21	21	2.7	33	1.84	0.40
Grades 2/3 mathematics	2012	2019	54	21	21	3.4	46	0.62	0.34
End of primary reading	2006	2019	78	66	40	15.3	20	-0.68	-0.35
End of primary mathematics	2003	2019	82	55	29	9.9	9	0.02	0.08
Lower secondary reading	2014	2019	28	16	0	0.0			
Lower secondary mathematics	2003	2019	30	18	2	0.9	15	1.26	0.67
Primary aged not out-of-school	2010	2019	73	72		50.8	79	2.3	
Primary completion	2010	2019	98	98		92.1	50	2.7	
Lower middle income countries									
Grades 2/3 reading	2006	2019	72	7	7	1.4	38	1.92	0.25
Grades 2/3 mathematics	2006	2019	72	7	7	1.3	50	2.32	0.40
End of primary reading	2001	2019	78	23	23	8.3	51	1.55	0.32
End of primary mathematics	2003	2019	80	18	16	5.8	25	0.89	0.10
Lower secondary reading	2000	2018	59	14	14	8.2	52	0.61	0.04
Lower secondary mathematics	2000	2019	67	23	23	14.2	37	0.99	0.25
Primary aged not out-of-school	2010	2019	91	75		43.4	92	0.9	
Primary completion	2010	2019	99	99		98.9	82	1.5	
Upper middle income countries									
Grades 2/3 reading	2006	2019	82	28	28	19.1	65	-0.20	-0.17
Grades 2/3 mathematics	2006	2019	82	25	25	17.3	60	0.69	0.52
End of primary reading	2001	2019	42	35	35	23.6	55	-0.19	-0.25
End of primary mathematics	2003	2019	45	36	36	24.8	51	-0.52	-0.48
Lower secondary reading	2000	2018	91	38	38	32.2	55	0.18	0.15
Lower secondary mathematics	2000	2019	91	38	38	31.6	42	0.39	0.31
Primary aged not out-of-school	2010	2019	44	44		39.3	97	0.6	
Primary completion	2010	2019	99	99		90.2	95	1.1	
High income countries									
Grades 2/3 reading	2001	2019	7	7	7	2.9	89	0.29	0.12
Grades 2/3 mathematics	2003	2019	7	7	7	4.5	70	1.19	0.72
End of primary reading	2001	2019	84	82	82	59.0	92	0.09	0.06
End of primary mathematics	2003	2019	87	85	85	57.6	71	0.26	0.15
Lower secondary reading	2000	2018	90	86	86	79.6	81	-0.02	-0.02
Lower secondary mathematics	2000	2019	92	91	91	84.2	72	-0.08	-0.09
Primary aged not out-of-school	2010	2019	92	92		77.1	99	0.5	
Primary completion	2010	2019	41	34		26.1	99	0.1	

Table 16: Population-focused World Bank income group breakdown with end of primary details

	Year range		% of population with a usable trend	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
Low income countries						
End of primary reading	2006	2019	32	9	-0.35	-0.16
End of primary mathematics	2003	2019	21	4	-0.02	0.03
Lower middle income countries						
End of primary reading	2001	2019	23	43	1.51	0.34
End of primary mathematics	2003	2019	16	20	0.81	0.10
Upper middle income countries						
End of primary reading	2001	2019	34	50	-0.01	-0.12
End of primary mathematics	2003	2019	36	48	-0.40	-0.40
High income countries						
End of primary reading	2001	2019	75	94	0.15	0.12
End of primary mathematics	2003	2019	74	74	0.35	0.23

Appendix 2: Trends in GPE countries

The next three tables provide details for the group of 85 countries within the Global Partnership for Education (GPE).

Table 17: Global Partnership for Education statistics

	Countries	Population aged 0 to 14 (thous.)	% of children covered by the largest quarter of countries	Year range for 4.1.1 statistics		% of population with any 4.1.1 data point	% of population with any 4.1.1 trend	% of population with any participation trend
GPE countries	85	831,372	77	2000	2019	91	57	100

Table 18: Global Partnership for Education with details by statistic

	Year range		% of population with any data point	% of population with any trend	% of population with a usable trend	Trend coverage (%)	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
GPE countries									
Grades 2/3 reading	2006	2019	61	16	16	3.1	37	1.75	0.26
Grades 2/3 mathematics	2006	2019	61	16	16	3.3	48	1.51	0.36
End of primary reading	2001	2019	74	48	38	12.9	41	0.93	-0.04
End of primary mathematics	2003	2019	76	38	26	7.7	21	0.48	-0.19
Lower secondary reading	2000	2019	35	23	17	9.8	52	0.61	0.05
Lower secondary mathematics	2000	2019	36	25	19	9.8	42	1.13	0.14
Primary aged not out-of-school	2010	2019	81	61		45.0	85	1.5	
Primary completion	2010	2019	100	100		97.2	67	2.3	

Table 19: Population-focused Global Partnership for Education breakdown with end of primary details

	Year range		% of population with a usable trend	Average level within year range	Annual percentage point gain (continuous approach)	Annual percentage point gain (restricted approach)
GPE countries						
End of primary reading	2001	2019	36	34	1.07	0.07
End of primary mathematics	2003	2019	24	16	0.49	-0.12

Appendix 3: A closer look at 2014 to 2019 PASEC trends

Ten PASEC countries had results for two levels of primary schooling and two subjects in both 2014 and 2019, which allowed for the determination of trends over time for these countries. This appendix draws from PASEC microdata and metadata made available to the UIS by PASEC, the aim here being to confirm the reliability of the trends. The analysis presented here could be taken further, for instance by extending it to both subjects and both levels of the primary system. The current analysis focuses on reading at the end of primary.

UIS (2021a: 19), in comparing learner responses on electricity in the household with World Bank data on electricity, concluded that in at least one PASEC country the national sample in 2019 may not have been consistent with the 2014 sample, which in turn could have affected the reliability of the learning proficiency trend. Follow-up analysis into this matter concluded that the way electricity information was collected in the PASEC background questionnaires for the two years makes comparison difficult. However, this further analysis found that Burkina Faso and Chad display levels of electricity access in 2019 in the PASEC data which are two to three times higher than what is found in the World Bank data, even after out-of-school children have been taken into account. *Importantly, other countries among the ten examined here do not display these discrepancies, suggesting that on the whole the PASEC samples are nationally representative.* Triangulation of data to examine the correctness of samples is important, given that in the past sampling issues have distorted trends in testing programmes other than PASEC. Yet it should be kept in mind that this type of analysis is more suggestive than conclusive. The ideal is some external quality assurance of the actual sampling process through examination of, in particular, the school and enrolment data used to determine the sampling frame, and publication of the results of this quality assurance step.

Figure 22 illustrates item response theory (IRT) score distributions in 2014 and 2019[51], while **Figure 23** presents distributions for classical percentage scores out of 100, based on just 41 common items used in both 2014 and 2019[52]. The patterns across the two figures can be considered highly consistent, and indicative of a correct scaling process when raw item responses were converted to IRT scores. For instance, Burundi's curve moved to the right and became steeper, regardless of which of the two figures is used. Niger displays a decline in a peak of worse performing learners, in both figures. This analysis confirms that the scaling process was clean, something already suggested by details provided in the PASEC technical documentation. One advantage with the kind of representation shown below is that it can help explain to teachers, who tend to be familiar with classical scores but not IRT scores, that distributions look relatively similar.

[51] The first of the five plausible values was used. Using the other five produced patterns which were virtually indistinguishable from those seen in Figure 22.

[52] A 42nd common item, f62 in 2019 named f63 in 2014, was excluded from the analysis as it displayed learner responses which were inconsistent across the two years, where the inconsistency was not easy to explain.

Figure 19: IRT end of primary reading trends in PASEC

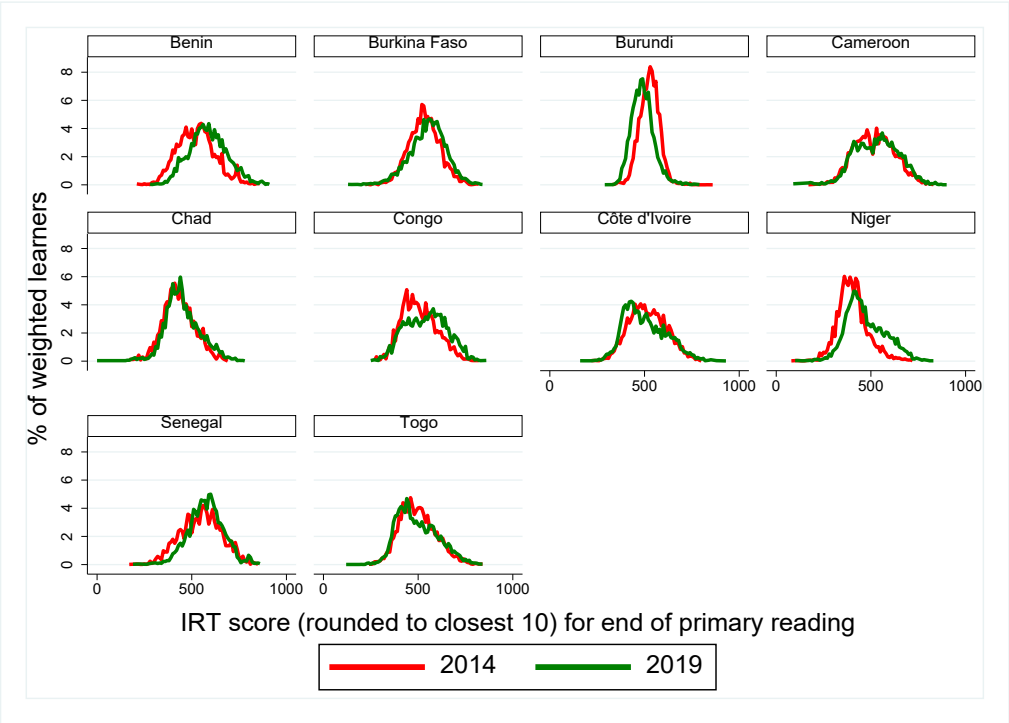
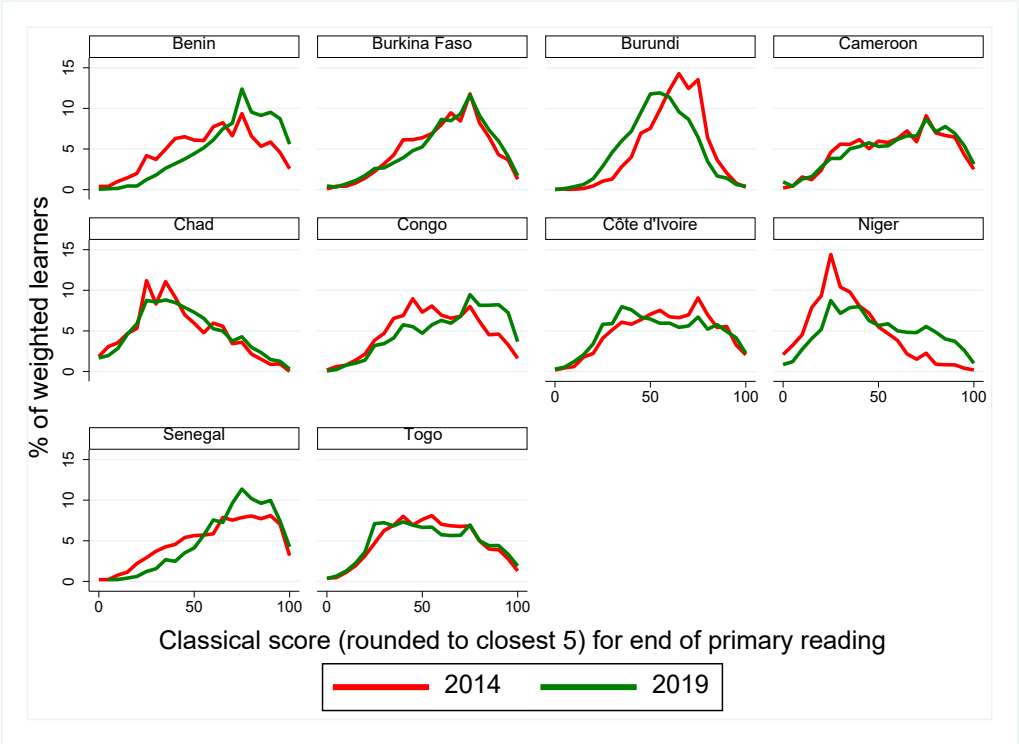


Figure 20: Classical end of primary reading trends based on anchor items in PASEC



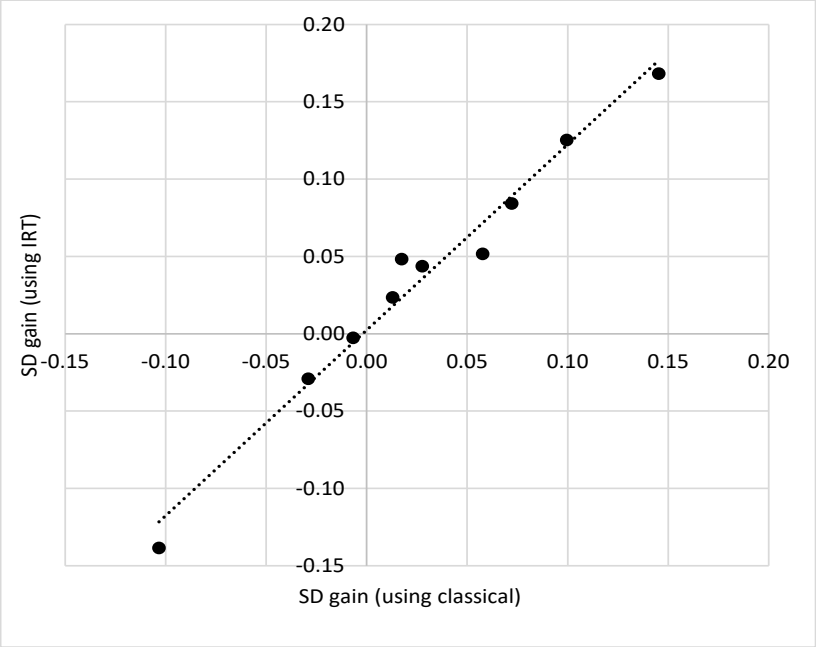
Though a similar analysis was not run for mathematics or for the lower primary level data, if the trends for end of primary reading appear consistent, it is very likely that other trends would also be consistent, as the methods and risks would be similar across the various PASEC trends.

Table 20 presents *annual* gains in terms of standard deviations, using IRT scores and classical scores. IRT means are the ones published by PASEC (CONFEMEN, 2020), while classical means were calculated using the microdata. Standard deviations for 2014 are all from the microdata. The gain statistics based on IRT means are highly correlated with those based on classical scores, which again confirms the absence of any serious problem in the scaling process, of the kind seen in a few other measures of national trends. The correlation coefficient across the two columns of annual gain statistics is 0.99. This high correlation is represented graphically in **Figure 24**.

Table 20: Classical and IRT end of primary reading trends compared

	IRT				Classical			
	2014 mean	2014 std. dev.	2019 mean	Annual gain in std. dev.	2014 mean	2014 std. dev.	2019 mean	Annual gain in std. dev.
Benin	523	99	586	.13	60	23	71	.10
Burkina Faso	532	82	552	.05	62	20	64	.02
Burundi	525	51	490	-.14	62	15	55	-.10
Cameroon	518	104	530	.02	60	24	61	.01
Chad	433	83	451	.04	41	21	44	.03
Congo	503	92	542	.08	57	22	65	.07
Côte d'Ivoire	517	97	503	-.03	59	22	56	-.03
Niger	404	80	471	.17	35	19	49	.15
Senegal	548	106	576	.05	64	23	71	.06
Togo	497	94	496	.00	55	22	54	-.01

Figure 21: Classical and IRT end of primary reading trends compared



UIS (2019) suggested that annual improvements greater than 0.6 standard deviations a year, using typical IRT measures, are rare. It moreover indicates that such high improvements can be expected in countries with a low point of departure in terms of learning proficiency levels. All the African countries of Table 20 can be considered to experience low levels of proficiency. The 2019 PASEC report warns that trends for reading in Chad and Niger over the 2014 to 2019 period should not be considered properly comparable, due to changes in the language used[53]. This leaves two countries from Table 20, Benin and Congo, with improvements that in the 'rare' category beyond a 0.06 annual standard deviation gain. These are two countries which according to the electricity access triangulation analysis referred to above display samples which are highly consistent across 2014 and 2019. Even an improvement as steep as Benin's 0.13 annual gain in standard deviations thus appears to be possible and plausible, though uncommon.

[53] CONFEMEN, 2020: 27, 220.

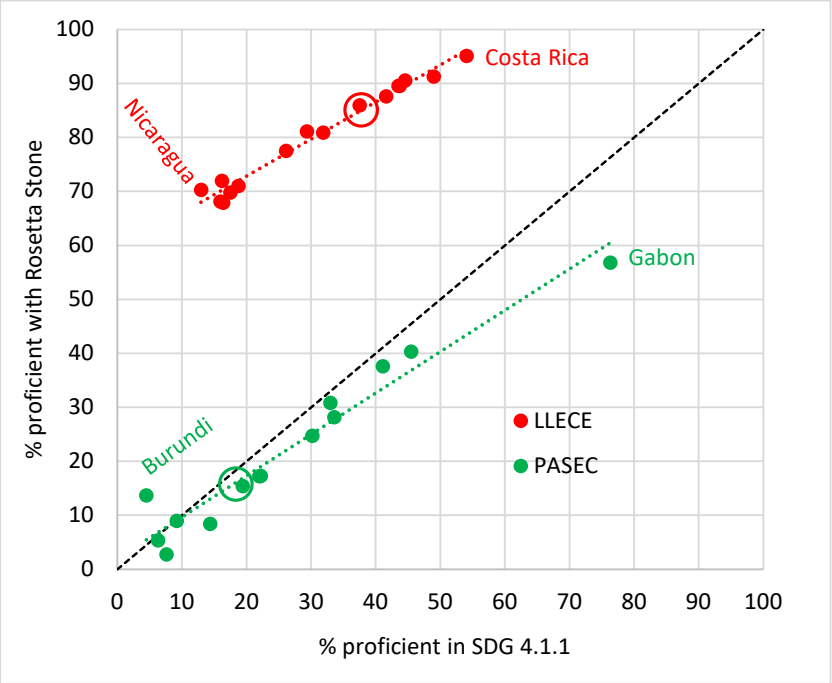
Appendix 4: Taking into account data issues in two regions

This appendix deals with issues in the Indicator 4.1.1 dataset used for the current report (see Section 2). The first issue is the high likelihood that proficiency statistics based on the LLECE programme are under-estimating actual proficiency in the Latin America and the Caribbean region by a large margin, due to LLECE's standards being particularly stringent. This matter, in turn, leads to a further question around the recent 2013 to 2019 trend in this region. The second issue is that the PILNA programme, which covers 15 countries in the Oceania region, is excluded from the 4.1.1 dataset as the results of individual countries within PILNA[54] are not published. However, results for PILNA participants as a group are published, and the impact of these on regional and global aggregate statistics can be calculated.

Figure 30 illustrates the findings of the Rosetta Stone project, which aimed to enhance the global comparability of 2019 statistics emerging from two programmes: PASEC and LLECE[55]. To illustrate, in the case of PASEC the project essentially arrived at specific scores on the PASEC Grade 6 reading scale that would be the equivalent of specific PIRLS reading scores, and PASEC Grade 6 mathematics scores that would be the equivalent of specific TIMSS mathematics scores. PIRLS and TIMSS are designed for Grade 4 learners, meaning the Rosetta Stone 'concordance tables' illustrate what PASEC Grade 6 learners would have achieved in a test designed for Grade 4. Despite the difference of two grades, percentage proficient statistics as published by PASEC are not very different to the proficiency statistics that would be obtained if the 'low international benchmark' of PIRLS or TIMSS were used. This is not surprising, as PIRLS and TIMSS were originally designed to reflect learning in largely developed countries. Specifically, where PASEC 2019 Grade 6 reading results point to 18% of learners in 14 countries in sub-Saharan Africa reaching at least a PASEC minimum proficiency level, the figure using the PIRLS low international benchmark is 16%. This is shown by the green hollow marker in Figure 25. Essentially the Rosetta Stone confirms that PASEC's standards are comparable to those applied in many other parts of the world, and those set by UNESCO as a global standard[56] (even if grades are different, as discussed above).

[54] Pacific Islands Literacy and Numeracy Assessment.
[55] UIS, 2022b; UIS, 2022c.
[56] UIS, 2020: 19.

Figure 30: Implications of Rosetta Stone for LLECE and PASEC



Note: Values refer to the reading proficiency of Grade 6 learners in 2019. Large hollow markers refer to the mean across countries weighted by the child population. The values against the vertical axis for LLECE were calculated as follows. First, PIRLS scores as a linear function of LLECE scores was found, using the concordance table and just the range within which LLECE countries are located (R squared was 0.997). Secondly, PIRLS proficiency percentages as a function of PIRLS mean scores using five developing countries and Mullis et al *(2017) was found (R squared of 0.979). The two functions were then used to convert published LLECE mean scores (UNESCO, 2021) to proficiency statistics using the low PIRLS benchmark. A similar approach was used for PASEC.*

The situation is very different with respect to LLECE. LLECE 2019 Grade 6 reading proficiency statistics produced a population-weighted mean across 16 countries of 38%, while the Rosetta Stone analysis resulted in a proficiency statistic of 85%. The latter figure can be considered more appropriate for reporting against SDG 4.1.1 than the first figure.

How would the use of the Rosetta Stone concordance table change the regional and global proficiency statistics seen in the third-last column of Table 9? Behind the Latin America and the Caribbean mean of 46% for end of primary reading seen in Table 9, all countries with statistics except for Trinidad and Tobago draw from LLECE. Trinidad and Tobago represents just 0.4% of the regional population with statistics, meaning the 46% seen in Table 9 becomes around 85% after the Rosetta Stone study is taken into account. There are no LLECE countries outside this region. Given that LLECE countries represent 6.9% of the global child population (regardless of whether countries have statistics or not), and given that the Table 9 proficiency statistic for the world outside Latin America and the Caribbean is 64%, based on a re-run of the code behind Table 9 with the region excluded, it can be concluded that the global proficiency level rises from the 60% seen for end of primary reading in Table 9 to 65%.

A further point regarding LLECE warrants attention, and this is not directly related to the Rosetta Stone initiative. SDG 4.1.1 statistics point to a substantial *decline* in Latin America and the Caribbean region at both primary levels between 2013 and 2019 (see Figure 17). However, this is driven by the fact that somewhat different methods were used for the proficiency statistics for the two years. The 2013 statistics are those initially published by LLECE, in 2014[57]. Subsequently, LLECE published proficiency statistics calculated in a manner that would make them comparable to published 2019 LLECE figures[58], which are those used as SDG 4.1.1 2019 values by the UIS. The subsequently published 2013 values are available in, for instance, Flotts *et al* (2015). The 2013 and 2019 comparable statistics from LLECE point to no gains, or small gains, but not to the substantial losses seen in Figure 17. To illustrate, the most recently published LLECE statistics point to programme-wide improvements between 2013 and 2019 of 35% to 38% with respect to the percentage proficient in Grade 6 reading (using population-weighted countries)[59].

In addition to comparability problems with respect to the 2013 and 2019 statistics published in UIS.Stat, LLECE has advised the UIS to avoid comparing 2006 results to 2013 results. This means that three adjustments to the March 2022 UIS.Stat dataset with respect to the Latin America and the Caribbean region, performed in the order specified here, can be justified: (1) removal of all 2006 LLECE values; (2) replacement of the 2013 values with those appearing in Flotts et al (2015); and (3) the adjustment of all LLECE Grade 6 statistics using the analysis presented in Figure 30[60].

If an adjusted UIS.Stat dataset, with the three abovementioned adjustments, is used to re-generate Table 9 above, all negative slopes for the two primary levels (last two columns of Table 9) in Latin America and the Caribbean become positive, the range being from 0.08 for the end of primary reading (restricted) to 3.82 for grades 2/3 mathematics (continuous). Globally, the adjustment improves the end of primary slopes somewhat, the most notable shift being from -0.04 to 0.44 for end of primary mathematics (continuous approach). For end of primary reading, the recalculated annual gains are 0.39 and -0.11 (continuous and restricted). Large apparent improvements at the global level for lower primary after the adjustment largely reflect the fact that the Latin America and the Caribbean region is the only region with extensive coverage of this level.

Turning to Oceania, the PILNA assessment programme has as its scope 15 island countries in the Pacific area encompassing 0.2% of the world's children, a figure that declines to 0.04% if the largest PILNA country, Papua New Guinea, is excluded. The 15 island countries encompass 40% of the Oceania region's child population (9% if Papua New Guinea is excluded). Four PILNA countries do appear in the UIS dataset of SDG proficiency statistics given the presence of alternative data sources. The four countries are: Kiribati; Samoa; Solomon Islands; and Tonga. However, none of these four countries have values for more than one year in the dataset, so trends are not discernible. To illustrate levels of proficiency, in Grades 2 to 3 (a level for which SDG statistics are available for the four countries) 55% of learners are proficient in reading. However, this ranges from 12% in Samoa to 71%

[57] UNESCO, 2014.
[58] See UNESCO, 2021.
[59] Flotts *et al*, 2015; UNESCO, 2021.
[60] The conversion from an x proficiency level in reading to y proficiency drawing from the Rosetta analysis used the following: $y = 59.12 + (0.6862 \times x)$. For mathematics, the following applied: $y = 59.90 + (0.9569 \times x)$.

in Solomon Islands. A 2019 report[61] on proficiency in the PILNA group of countries points for instance to Grade 6 literacy proficiency improving from 49% to 63% between 2012 to 2018.

PILNA countries cover sufficient children for the abovementioned statistics to make a difference to the regional Oceania statistics from Table 9, but not enough to children to make any noticeable difference to the global statistics represented in the same table. To illustrate the effects on Oceania, incorporating PILNA results lower the level of proficiency for end of primary reading from the 91% seen in Table 9 to 77%, while improving the annual gain marginally from -0.01 to zero.

[61] Pacific Community, 2019.